Journal	**: Journal of Computer Science and Information Technology (JCSIT)**
Publisher	**: Yayasan Nuraini Ibrahim Mandiri**
ISSN	**: 3031-8467**
Volume / Issue / Year/ Page	**: Vol. 3 No. 4 (2026): September/ Hal 542-552**

KLASIFIKASI KELAYAKAN SAHAM JANGKA PANJANG MENGGUNAKAN RANDOM FOREST BERBASIS RASIO FUNDAMENTAL EMITEN BEI

Afil Anugrah[1*], Rizki Yusliana Bakti[2], Muhammad Faisal[3]
Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Makassar
[1]105841113522@student.unismuh.ac.id, [2]rizkiyusliana@unismuh.ac.id, [3]muhfaisal@unismuh.ac.id
* Coresponding Author

Info Article

Pengajuan : 09-08-2026
Peninjauan : 20-08-2026
Revisi : 30-08-2026
Diterima : 10-09-2026
Diterbitkan : 28-09-2026

Keywords:
Random Forest, Rasio Fundamental, Penyaringan Saham, Bursa Efek Indonesia, Kepentingan Fitur

Abstrak

Investasi saham jangka panjang memerlukan penyaringan yang mempertimbangkan kondisi fundamental perusahaan secara sistematis. Penelitian ini bertujuan membangun model klasifikasi kelayakan saham menggunakan Random Forest berbasis Price Earnings Ratio (PER), Price to Book Value (PBV), Return on Equity (ROE), Debt to Equity Ratio (DER), dan Earnings per Share (EPS) pada emiten Bursa Efek Indonesia. Dataset terdiri atas 375 perusahaan dengan data fundamental periode 2025 yang diperoleh menggunakan yfinance. Label dibentuk melalui rule-based scoring dengan kriteria PER < 25, PBV < 3, ROE > 15, DER < 1, dan EPS > 0; skor ≥ 3 dikategorikan Layak dan skor < 3 Tidak Layak. Data dibagi 80:20 menjadi 300 data training dan 75 data testing. Random Forest menggunakan 100 Decision Tree dan dievaluasi dengan Accuracy, Precision, Recall, F1-Score, dan Confusion Matrix. Hasil pengujian menunjukkan Accuracy 97,33%, Precision 97,92%, Recall 97,92%, dan F1-Score 97,92%, dengan 73 dari 75 data uji diklasifikasikan benar. Feature Importance menunjukkan PER 0,47, PBV 0,24, EPS 0,13, ROE 0,11, dan DER 0,05. Hasil tersebut menunjukkan konsistensi model dalam mereplikasi label berbasis aturan, bukan bukti kemampuan memprediksi return saham masa depan karena label dan fitur berasal dari indikator yang sama.

How to Cite : Anugrah, A., Bakti, R. Y., & Faisal, M. (2026). Klasifikasi kelayakan saham jangka panjang menggunakan Random Forest berbasis rasio fundamental emiten BEI. Journal of Computer Science and Information Technology (JCSIT), 3(4), 531–541. https://doi.org/10.70248/jcsit.v3i4.4453

DOI : https://doi.org/10.70248/jcsit.v3i4.4453
This is an open access article under https://creativecommons.org/licenses/by/4.0/
Copyright© 2026 by Author. Published by Yayasan Nuraini Ibrahim Mandiri

PENDAHULUAN

Pasar modal merupakan bagian penting dari sistem perekonomian karena menyediakan mekanisme penghimpunan dana dan mempertemukan pihak yang membutuhkan modal dengan pihak yang memiliki dana untuk diinvestasikan. Di Indonesia, Bursa Efek Indonesia (BEI) menjadi salah satu pusat aktivitas perdagangan saham yang menyediakan informasi dan mekanisme perdagangan bagi emiten dan investor. Investasi saham menawarkan peluang memperoleh imbal hasil, tetapi keputusan investasi tetap menghadapi ketidakpastian karena kondisi perusahaan, industri, makroekonomi, dan pasar dapat berubah dari waktu ke waktu. Oleh sebab itu, proses penyaringan saham membutuhkan informasi yang dapat digunakan secara sistematis untuk menilai karakteristik fundamental perusahaan (Alam et al., 2025; Dewi Lubis et al., 2024).

Dalam konteks investasi jangka panjang, analisis fundamental digunakan untuk melihat kondisi ekonomi dan keuangan perusahaan melalui sejumlah indikator. PER menggambarkan hubungan antara harga pasar saham dan laba per saham, sedangkan

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

PBV membandingkan nilai pasar dengan nilai buku. ROE memberikan informasi mengenai kemampuan perusahaan menghasilkan laba terhadap ekuitas, DER menggambarkan struktur pendanaan dan tingkat penggunaan utang terhadap ekuitas, sedangkan EPS menunjukkan laba yang tersedia untuk setiap lembar saham. Kelima rasio tersebut tidak berdiri sendiri karena masing-masing merepresentasikan aspek valuasi, profitabilitas, maupun struktur keuangan yang berbeda. Penelitian Ferniawan et al. (2024), misalnya, menggunakan EPS, PER, PBV, dan DER dalam menguji keterkaitannya dengan return saham perusahaan yang terdaftar di BEI.

Perkembangan machine learning memberikan alternatif untuk mengolah data fundamental dalam jumlah lebih besar dan membangun model klasifikasi secara otomatis. Random Forest merupakan metode ensemble yang membangun banyak Decision Tree melalui bootstrap sampling dan pemilihan fitur secara acak, kemudian menggabungkan prediksi pohon menggunakan majority voting. Karakteristik tersebut memungkinkan Random Forest membentuk batas keputusan yang tidak harus linier dan menyediakan ukuran kepentingan fitur. Penelitian terbaru mengenai saham menunjukkan bahwa Random Forest telah digunakan pada pemodelan harga dan penggabungan indikator teknikal maupun fundamental, sedangkan studi lain menggunakan Random Forest untuk memodelkan data historis saham Bank BRI (Mahfuzi & Munawaroh, 2026; Rasyid et al., 2025).

Namun, terdapat perbedaan mendasar antara prediksi harga saham dan klasifikasi kelayakan berbasis rasio fundamental. Prediksi harga umumnya menggunakan harga historis, volume, atau indikator teknikal sebagai masukan dan mengukur kesalahan terhadap nilai harga aktual. Penelitian ini menggunakan klasifikasi dua kelas, yaitu Layak dan Tidak Layak, dengan label yang dibentuk melalui rule-based scoring dari lima rasio fundamental. Dengan demikian, model Random Forest pada penelitian ini tidak diposisikan sebagai model prediksi return masa depan, tetapi sebagai model yang diuji untuk mempelajari dan mereplikasi keputusan screening berbasis aturan.

Persoalan metodologis tersebut perlu dinyatakan secara eksplisit karena label target dan variabel input berasal dari sumber indikator yang sama. Kondisi ini dapat menghasilkan akurasi tinggi tanpa menunjukkan bahwa model menemukan pola laten baru di pasar. Jika suatu perusahaan diberi label Layak karena memenuhi sebagian besar ambang PER, PBV, ROE, DER, dan EPS, maka Random Forest yang menggunakan kelima rasio tersebut memang memiliki informasi yang sangat dekat dengan mekanisme pembentukan label. Fenomena tersebut dapat disebut sebagai rule-replication bias dalam konteks penelitian ini. Oleh karena itu, akurasi 97,33% harus ditafsirkan sebagai tingkat konsistensi model terhadap label rule-based, bukan sebagai bukti bahwa model dapat memprediksi kinerja saham di masa depan.

Meski demikian, pengujian Random Forest tetap memiliki nilai metodologis untuk melihat apakah kumpulan Decision Tree mampu merepresentasikan aturan dengan stabil pada data yang tidak digunakan selama pelatihan. Selain itu, Feature Importance dapat digunakan untuk melihat variabel mana yang paling sering dimanfaatkan model dalam membentuk pemisahan kelas. Informasi ini dapat membantu menjelaskan struktur keputusan model, walaupun tidak boleh disamakan dengan pengaruh kausal terhadap return saham.

Penelitian terdahulu yang dirujuk dalam naskah banyak membahas prediksi harga, indikator teknikal, atau perbandingan model machine learning (Apridius & Lasut, 2025;

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

Suyatna et al., 2025). Studi lain juga menunjukkan bahwa faktor fundamental dan makroekonomi dapat memberikan informasi tambahan dalam pemodelan saham pada horizon yang lebih panjang (IIMB Management Review, 2025). Ruang penelitian yang digunakan dalam artikel ini terletak pada penggunaan lima rasio fundamental untuk membentuk label screening dan kemudian menguji kemampuan Random Forest dalam mereplikasi label tersebut serta menjelaskan kontribusi relatif setiap rasio.

Berdasarkan latar belakang tersebut, penelitian ini memiliki dua fokus utama. Pertama, membangun model klasifikasi kelayakan saham jangka panjang menggunakan PER, PBV, ROE, DER, dan EPS dengan algoritma Random Forest. Kedua, mengevaluasi kinerja model dan mengidentifikasi kontribusi relatif variabel melalui Feature Importance, sekaligus membahas batas interpretasi hasil karena label dibentuk dari aturan yang menggunakan fitur yang sama. Dengan fokus tersebut, penelitian tidak dimaksudkan untuk memberikan rekomendasi investasi individual, melainkan untuk mengevaluasi pendekatan komputasional sebagai alat screening awal.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis machine learning. Objek penelitian adalah perusahaan yang terdaftar di Bursa Efek Indonesia dan memiliki data fundamental yang dapat diperoleh melalui library Python yfinance. Periode pengamatan yang digunakan adalah tahun 2025. Setelah proses pengumpulan dan pemeriksaan data, diperoleh 375 perusahaan yang memiliki nilai pada lima variabel fundamental, yaitu PER, PBV, ROE, DER, dan EPS. Data tersebut digunakan sebagai unit observasi sehingga satu baris merepresentasikan satu perusahaan pada periode pengamatan.

Pengumpulan data dilakukan dengan memanfaatkan yfinance untuk memperoleh informasi fundamental yang tersedia pada masing-masing emiten. Data kemudian disusun ke dalam dataset terstruktur. Tahap preprocessing meliputi pemeriksaan tipe data, missing value, duplikasi, dan nilai yang tidak dapat digunakan dalam perhitungan rasio. Hasil pemeriksaan pada dataset penelitian menunjukkan tidak ditemukan missing value pada lima variabel dan tidak ditemukan data duplikat pada 375 observasi. Tahapan ini dilakukan sebelum pelabelan dan pembagian data agar dataset yang digunakan dalam pemodelan berada dalam kondisi konsisten.

Tabel 1. Variabel yang digunakan dalam klasifikasi

No.	Variabel	Jenis Data	Keterangan
1	PER	Numerik	Price Earnings Ratio
2	PBV	Numerik	Price to Book Value
3	ROE	Numerik	Return on Equity
4	DER	Numerik	Debt to Equity Ratio
5	EPS	Numerik	Earnings per Share

Pelabelan Rule-Based Scoring

Pelabelan dilakukan menggunakan sistem rule-based scoring. Setiap indikator memperoleh skor 1 apabila memenuhi kriteria dan skor 0 apabila tidak memenuhi kriteria. Ambang yang digunakan adalah PER < 25, PBV < 3, ROE > 15, DER < 1, dan EPS > 0. Total skor dihitung dengan menjumlahkan skor kelima indikator. Perusahaan dengan total skor ≥ 3 diberi label Layak, sedangkan perusahaan dengan skor < 3 diberi label Tidak Layak.

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

Ambang tersebut perlu dipahami sebagai operational threshold penelitian, bukan sebagai standar universal yang secara tunggal menentukan apakah suatu saham layak dibeli. Literatur menunjukkan bahwa PER, PBV, EPS, dan DER digunakan dalam analisis fundamental dan memiliki hubungan dengan karakteristik return atau valuasi saham, tetapi nilai ambang dapat berbeda menurut sektor, periode, dan tujuan analisis. Karena itu, penelitian ini menggunakan ambang tersebut sebagai aturan operasional untuk membentuk target klasifikasi, sementara validasi berbasis kinerja pasar aktual masih diperlukan untuk menilai kelayakan ekonominya (Ferniawan et al., 2024; Sihombing et al., 2026).

Tabel 2. Kriteria rule-based scoring

Variabel	Kriteria	Skor
PER	PER < 25	1
PBV	PBV < 3	1
ROE	ROE > 15	1
DER	DER < 1	1
EPS	EPS > 0	1

Pembagian Data dan Ketidakseimbangan Kelas

Setelah pelabelan, diperoleh 241 data Layak dan 134 data Tidak Layak. Proporsi tersebut masing-masing sebesar 64,27% dan 35,73%. Distribusi ini menunjukkan ketidakseimbangan kelas yang bersifat moderat. Eksperimen utama tidak menerapkan SMOTE maupun class weight karena penelitian diarahkan untuk mengukur performa dasar model pada distribusi observasi yang tersedia. Namun, kondisi tersebut tetap dipertimbangkan dalam interpretasi melalui Precision, Recall, F1-Score, dan Confusion Matrix, bukan hanya Accuracy.

Dataset dibagi menggunakan hold-out validation dengan rasio 80:20. Sebanyak 300 observasi digunakan sebagai data training dan 75 observasi sebagai data testing. Pembagian ini menjaga pemisahan antara data yang digunakan untuk membangun model dan data yang digunakan untuk evaluasi akhir. Keterbatasan pendekatan hold-out adalah hasil dapat dipengaruhi oleh komposisi satu pembagian data. Oleh sebab itu, penerapan Stratified K-Fold Cross-Validation pada eksperimen lanjutan direkomendasikan untuk mengukur stabilitas performa model pada beberapa pembagian data.

Tabel 3. Pembagian data latih dan data uji

Jenis Data	Persentase	Jumlah Data
Data Latih	80%	300
Data Uji	20%	75
Total	100%	375

Konfigurasi Random Forest

Model klasifikasi dibangun menggunakan RandomForestClassifier dari pustaka Scikit-learn. Model utama menggunakan n_estimators=100 sehingga terdiri atas 100 Decision Tree. Parameter lain menggunakan konfigurasi default Scikit-learn, yaitu criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, max_features='sqrt', bootstrap=True, dan class_weight=None. Konfigurasi tersebut dipertahankan agar eksperimen dapat direplikasi dan tidak memasukkan optimasi hyperparameter yang tidak terdokumentasi dalam eksperimen awal.

Random Forest bekerja dengan membentuk beberapa pohon dari sampel bootstrap

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

dan memilih subset fitur secara acak ketika membangun percabangan. Setiap pohon dapat menghasilkan keputusan yang berbeda karena data dan atribut yang digunakan pada pembentukan pohon tidak sepenuhnya sama. Prediksi akhir ditentukan berdasarkan majority voting. Mekanisme ensemble tersebut digunakan untuk mengevaluasi apakah aturan yang dibentuk dari lima rasio dapat direplikasi secara konsisten pada data testing.

Tabel 4. Konfigurasi model Random Forest

Parameter	Nilai
n_estimators	100
criterion	gini
max_depth	None
min_samples_split	2
min_samples_leaf	1
max_features	sqrt
bootstrap	True
class_weight	None

Evaluasi Model

Evaluasi dilakukan menggunakan Accuracy, Precision, Recall, F1-Score, dan Confusion Matrix. Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data testing. Precision mengukur ketepatan prediksi suatu kelas, Recall mengukur kemampuan model menemukan data aktual pada kelas tersebut, sedangkan F1-Score merupakan gabungan antara Precision dan Recall. Confusion Matrix digunakan untuk melihat distribusi True Positive, True Negative, False Positive, dan False Negative.

Selain evaluasi hold-out, Stratified K-Fold Cross-Validation merupakan rancangan validasi yang disarankan untuk eksperimen berikutnya karena dapat mempertahankan proporsi kelas pada setiap fold. Namun, nilai rata-rata dan simpangan baku cross-validation tidak dicantumkan sebagai hasil penelitian ini karena eksperimen yang terdokumentasi hanya menyediakan hasil hold-out 80:20. Dengan demikian, tidak ada angka cross-validation yang direka untuk menggantikan hasil aktual.

HASIL PENELITIAN DAN PEMBAHASAN

Dataset akhir terdiri atas 375 perusahaan dengan lima variabel fundamental. Pemeriksaan preprocessing menunjukkan tidak terdapat missing value dan duplikasi pada variabel yang digunakan. Proses rule-based scoring menghasilkan 241 perusahaan dengan label Layak dan 134 perusahaan dengan label Tidak Layak. Distribusi tersebut menjadi dasar dalam pembagian data dan interpretasi hasil evaluasi.

Tabel 5. Distribusi hasil pelabelan

Label	Jumlah Data	Persentase
Layak	241	64,27%
Tidak Layak	134	35,73%
Total	375	100%

Distribusi kelas menunjukkan bahwa kelas Layak lebih besar daripada kelas Tidak Layak dengan selisih 107 observasi. Perbedaan tersebut tidak ekstrem, tetapi dapat memengaruhi Accuracy apabila model lebih sering mengikuti kelas mayoritas. Oleh karena itu, hasil penelitian tidak hanya dibaca melalui Accuracy, tetapi juga melalui Precision, Recall, F1-Score dan pola kesalahan pada Confusion Matrix. Tidak digunakannya SMOTE atau class weight merupakan bagian dari desain eksperimen

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

dasar dan sekaligus menjadi keterbatasan yang dapat diuji lebih lanjut.

Setelah pembagian data, Random Forest dilatih menggunakan 300 observasi. Setiap pohon menggunakan mekanisme bootstrap sampling dan random feature selection, kemudian 100 hasil pohon digabungkan melalui majority voting. Data testing yang berjumlah 75 observasi tidak digunakan pada proses fitting sehingga dapat digunakan untuk mengukur konsistensi model terhadap data yang tidak dilihat selama pelatihan.

Tabel 6. Hasil evaluasi model Random Forest

Metrik	Nilai
Accuracy	0,9733 (97,33%)
Precision	0,9792 (97,92%)
Recall	0,9792 (97,92%)
F1-Score	0,9792 (97,92%)

Hasil evaluasi menunjukkan Accuracy sebesar 0,9733 atau 97,33%. Nilai tersebut konsisten dengan Confusion Matrix karena terdapat 73 prediksi benar dari 75 data testing. Precision, Recall, dan F1-Score masing-masing sebesar 0,9792 atau 97,92%. Secara deskriptif, angka tersebut menunjukkan bahwa model mampu mereplikasi sebagian besar label rule-based pada data testing.

Tabel 7. Confusion Matrix Random Forest

Aktual / Prediksi	Prediksi Layak	Prediksi Tidak Layak
Layak	47	1
Tidak Layak	1	26

Confusion Matrix memperlihatkan 47 data Layak diprediksi benar sebagai Layak dan 26 data Tidak Layak diprediksi benar sebagai Tidak Layak. Terdapat satu data Layak yang diprediksi sebagai Tidak Layak dan satu data Tidak Layak yang diprediksi sebagai Layak. Dengan demikian, jumlah prediksi benar adalah 47 + 26 = 73 data, sedangkan dua data mengalami kesalahan. Accuracy dapat diverifikasi melalui perhitungan (47 + 26) / 75 = 0,9733.

Performa yang tinggi tersebut tidak dapat langsung ditafsirkan sebagai bukti kemampuan model dalam memprediksi saham yang akan menghasilkan return tinggi. Target penelitian dibentuk dari aturan yang menggunakan PER, PBV, ROE, DER, dan EPS, sedangkan kelima rasio yang sama digunakan sebagai fitur masukan Random Forest. Model dengan demikian memperoleh informasi yang secara langsung berkaitan dengan mekanisme pembentukan label. Akurasi tinggi sebagian dapat dijelaskan oleh kemampuan model mereplikasi aturan tersebut.

Fenomena ini merupakan batasan penting yang disebut sebagai rule-replication bias. Dalam klasifikasi berbasis rule-generated labels, keberhasilan model terhadap label buatan tidak sama dengan keberhasilan memprediksi fenomena eksternal yang belum diketahui. Jika tujuan penelitian adalah memprediksi performa investasi, target seharusnya berasal dari kinerja masa depan, misalnya cumulative return, excess return, atau ukuran kinerja portofolio pada horizon tertentu. Penggunaan label tersebut akan memisahkan informasi pada saat pembentukan fitur dari outcome yang terjadi setelah periode observasi.

Karena itu, hasil penelitian ini sebaiknya diposisikan sebagai evaluasi konsistensi model screening berbasis aturan. Random Forest memberikan representasi komputasional yang dapat digunakan untuk memproses data baru dengan prosedur yang sama, tetapi penelitian belum membuktikan bahwa model tersebut lebih unggul daripada kaskade IF-THEN sederhana. Perbandingan langsung dengan rule engine,

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

model statistik, atau classifier lain pada label yang sama dapat menjadi eksperimen tambahan, sedangkan validasi menggunakan return masa depan merupakan langkah yang lebih penting untuk menguji relevansi ekonominya.

Analisis Feature Importance

Feature Importance digunakan untuk melihat kontribusi relatif setiap variabel dalam pembentukan keputusan Random Forest. Nilai yang diperoleh merupakan ukuran penggunaan fitur dalam proses pembentukan Decision Tree dan tidak boleh ditafsirkan sebagai hubungan sebab-akibat. Hasil Feature Importance pada dataset penelitian disajikan pada Tabel 8.

Tabel 8. Nilai Feature Importance

No.	Variabel	Feature Importance
1	PER (Price Earnings Ratio)	0,47
2	PBV (Price to Book Value)	0,24
3	EPS (Earnings per Share)	0,13
4	ROE (Return on Equity)	0,11
5	DER (Debt to Equity Ratio)	0,05

PER memiliki Feature Importance sebesar 0,47, sedangkan PBV sebesar 0,24. Dua rasio tersebut merupakan indikator valuasi yang secara langsung menghubungkan harga pasar dengan laba atau nilai buku. Pada penelitian ini, keduanya juga memiliki ambang eksplisit dalam rule-based scoring. Oleh sebab itu, dominasi PER dan PBV dapat muncul karena posisi observasi terhadap ambang tersebut menghasilkan pemisahan kelas yang kuat pada Decision Tree.

EPS memiliki nilai 0,13 dan ROE 0,11. Kedua variabel tersebut merepresentasikan aspek profitabilitas dan tetap digunakan dalam aturan pembentukan label. DER memiliki nilai 0,05 dan menjadi variabel dengan kontribusi relatif paling kecil. Perbedaan tersebut tidak berarti DER tidak penting secara finansial. Nilai Feature Importance dipengaruhi oleh distribusi data, variasi nilai di sekitar ambang, korelasi antarfitur, serta cara Decision Tree memilih split. Karena itu, Feature Importance pada penelitian ini lebih tepat dipahami sebagai kontribusi komputasional terhadap label penelitian.

Dominasi PER dan PBV juga perlu dikaitkan dengan konteks literatur fundamental. Penelitian Ferniawan et al. (2024) menunjukkan bahwa EPS, PER, PBV, dan DER dapat digunakan sebagai variabel dalam analisis return saham. Sihombing et al. (2026) juga membahas penggunaan ROE, PER, PBV, dan DER dalam pengambilan keputusan investasi saham sektor perbankan. Namun, penelitian tersebut tidak menjadi dasar untuk menyimpulkan bahwa PER dan PBV selalu lebih menentukan return daripada rasio lainnya. Dalam artikel ini, dominasi PER dan PBV terutama terkait dengan struktur label yang dibentuk dari lima rasio yang sama.

Temuan Feature Importance juga memperkuat alasan mengapa model perlu dibaca bersama dengan desain pelabelan. Apabila target ditentukan dari aturan yang memberi kontribusi satu skor kepada setiap rasio, maka fitur yang berada dekat dengan batas keputusan dapat memberikan pemisahan yang lebih informatif bagi pohon. Dengan demikian, nilai 0,47 pada PER tidak dapat diterjemahkan sebagai '47% pengaruh terhadap return' dan nilai 0,24 pada PBV tidak berarti '24% peluang saham meningkat'. Interpretasi yang tepat adalah bahwa kedua fitur tersebut memiliki kontribusi relatif lebih besar dalam pembentukan keputusan model pada dataset dan label yang digunakan.

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

Perbandingan dengan Penelitian Terdahulu dan Keunikan Penelitian

Penelitian Random Forest pada saham memiliki fokus yang beragam. Mahfuzi dan Munawaroh (2026) menggunakan data historis harga Bank BRI dan mengevaluasi model dengan MAE dan RMSE, sedangkan Apridius dan Lasut (2025) menggunakan indikator Exponential Moving Average untuk memprediksi harga. Rasyid et al. (2025) menggabungkan analisis fundamental, teknikal, dan bandarmologi untuk prediksi harga saham syariah. Studi ensemble pada saham perbankan juga menunjukkan bahwa kombinasi faktor teknikal, fundamental, dan makroekonomi dapat digunakan untuk membentuk model prediksi pada beberapa horizon waktu (IIMB Management Review, 2025).

Keunikan penelitian ini bukan pada klaim bahwa Random Forest menghasilkan prediksi return masa depan, tetapi pada pembentukan masalah sebagai klasifikasi screening dua kelas berbasis rasio fundamental. Lima indikator digunakan secara eksplisit untuk membangun label, kemudian Random Forest diuji untuk melihat konsistensi reproduksi label tersebut dan memberikan informasi Feature Importance. Framing ini membuat batas kontribusi penelitian lebih jelas: hasil utama merupakan bukti kemampuan model mereplikasi aturan pada dataset 2025, bukan bukti superioritas investasi.

Dari sisi praktis, model dapat diposisikan sebagai alat bantu screening awal untuk mengurangi pekerjaan penyaringan manual ketika jumlah saham yang perlu diperiksa cukup besar. Namun, hasil model harus dipadukan dengan analisis sektor, kondisi perusahaan, perkembangan laporan keuangan, kondisi pasar, dan faktor ekonomi lain. Model juga tidak dirancang untuk menggantikan analisis investor atau menjadi satu-satunya dasar keputusan investasi.

Keterbatasan Penelitian dan Implikasi Pengembangan

Penelitian memiliki beberapa keterbatasan. Pertama, dataset hanya menggunakan satu periode, yaitu 2025, sehingga variasi kondisi ekonomi dan pasar antarwaktu belum terwakili. Kedua, jumlah observasi 375 perusahaan relatif terbatas apabila dibandingkan dengan kebutuhan model yang ingin digeneralisasikan lintas periode dan sektor. Ketiga, label berasal dari rule-based scoring menggunakan fitur yang sama dengan input model sehingga terdapat risiko rule-replication bias. Keempat, validasi utama menggunakan satu hold-out split 80:20 sehingga stabilitas hasil terhadap perubahan pembagian data belum diketahui.

Kelima, threshold PER < 25, PBV < 3, ROE > 15, DER < 1, dan EPS > 0 merupakan aturan operasional penelitian. Nilai tersebut dapat memberikan hasil berbeda jika diterapkan pada sektor yang memiliki karakteristik rasio yang berbeda. Karena itu, penggunaan threshold yang sama untuk seluruh emiten perlu dievaluasi lebih lanjut melalui analisis sektoral atau pendekatan threshold yang mempertimbangkan distribusi industri.

Pengembangan berikutnya dapat menggunakan panel data multi-tahun agar model memperoleh variasi temporal dan dapat diuji pada periode yang berbeda. Label dapat dibentuk menggunakan outcome aktual, misalnya return 6 bulan atau 12 bulan setelah tanggal observasi, excess return terhadap indeks acuan, atau ukuran kinerja portofolio. Dengan desain tersebut, fitur fundamental pada periode t digunakan untuk menjelaskan outcome pada periode t+1 sehingga hubungan temporal lebih jelas dan risiko

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

kebocoran target dapat dikurangi.

Validasi juga dapat diperkuat dengan Stratified K-Fold Cross-Validation dan hyperparameter tuning. Hasil setiap fold sebaiknya dilaporkan sebagai mean dan standard deviation untuk Accuracy, Precision, Recall, dan F1-Score. Penelitian lanjutan juga dapat membandingkan Random Forest dengan rule engine deterministik, Logistic Regression, Support Vector Machine, Gradient Boosting, dan model ensemble lain pada label yang sama. Jika tujuan akhirnya adalah menilai kinerja investasi, perbandingan sebaiknya menggunakan metrik pasar dan bukan hanya metrik klasifikasi.

Tabel 9. Ringkasan tahapan penelitian

Tahap	Kegiatan	Output
1	Pengumpulan data melalui yfinance	Data fundamental 375 perusahaan
2	Preprocessing	Data bersih tanpa missing value dan duplikasi
3	Rule-based scoring	Label Layak dan Tidak Layak
4	Pembagian data 80:20	300 training dan 75 testing
5	Pelatihan Random Forest	Model 100 Decision Tree
6	Evaluasi	Accuracy, Precision, Recall, F1-Score, Confusion Matrix
7	Interpretasi	Feature Importance dan analisis keterbatasan

KESIMPULAN

Penelitian ini menerapkan Random Forest untuk mengklasifikasikan kelayakan saham jangka panjang berdasarkan lima rasio fundamental, yaitu PER, PBV, ROE, DER, dan EPS. Dataset terdiri atas 375 perusahaan periode 2025 dan diberi label menggunakan rule-based scoring dengan kategori Layak untuk total skor ≥ 3 dan Tidak Layak untuk skor < 3. Pembagian data menggunakan hold-out validation 80:20 menghasilkan 300 data training dan 75 data testing.

Hasil evaluasi pada data testing menunjukkan Accuracy sebesar 97,33%, Precision 97,92%, Recall 97,92%, dan F1-Score 97,92%. Confusion Matrix menunjukkan 47 data Layak dan 26 data Tidak Layak diklasifikasikan dengan benar, sedangkan satu data pada masing-masing kelas mengalami kesalahan klasifikasi. Feature Importance menunjukkan kontribusi relatif PER sebesar 0,47, PBV 0,24, EPS 0,13, ROE 0,11, dan DER 0,05.

Temuan tersebut menunjukkan bahwa Random Forest mampu mereplikasi label rule-based dengan tingkat konsistensi yang tinggi pada data testing. Namun, tingginya metrik klasifikasi tidak dapat digunakan untuk menyimpulkan kemampuan prediksi return saham masa depan karena label dibentuk dari indikator yang sama dengan fitur masukan. Dengan demikian, keterbatasan utama penelitian terletak pada rule-replication bias, penggunaan data satu periode, ukuran sampel, dan validasi hold-out. Penelitian selanjutnya disarankan menggunakan panel data multi-tahun, Stratified K-Fold Cross-Validation, tuning hyperparameter, serta label berbasis kinerja pasar aktual pada periode setelah observasi agar kemampuan generalisasi dan relevansi ekonominya dapat diuji secara lebih kuat.

Pernyataan Mengenai Penggunaan Kecerdasan Buatan (AI)

AI digunakan dalam penyusunan naskah ini. Penulis menggunakan perangkat kecerdasan buatan semata-mata untuk membantu perbaikan bahasa, pemeriksaan tata

bahasa, serta penyuntingan demi kejelasan dan keterbacaan. Penulis tetap bertanggung jawab penuh atas akurasi, orisinalitas, integritas, dan konten akademis naskah tersebut. Seluruh konten yang dibantu oleh AI telah ditinjau dan diverifikasi oleh penulis sebelum naskah diserahkan.

Konflik Kepentingan

Penulis menyatakan tidak ada konflik kepentingan.

Kontribusi Penulis

Konseptualisasi, Afil Anugrah dan Rizki Yusliana Bakti; pengumpulan data, Afil Anugrah; pengolahan dan preprocessing data, Afil Anugrah; analisis formal dan pemodelan, Afil Anugrah dan Rizki Yusliana Bakti; validasi metodologi, Rizki Yusliana Bakti dan Muhammad Faisal; penulisan—penyusunan draf awal, Afil Anugrah; penulisan—peninjauan dan penyuntingan, Rizki Yusliana Bakti dan Muhammad Faisal; visualisasi, Afil Anugrah. Seluruh penulis telah membaca dan menyetujui versi akhir naskah ini.

Ucapan Terima Kasih

Penelitian ini didukung oleh Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Makassar. Penulis menyampaikan terima kasih atas dukungan akademik, fasilitas, dan sumber daya yang diberikan selama proses penelitian dan penyusunan naskah.

DAFTAR PUSTAKA

Alam, D., Kurniawan, A., & Hanggraeni, D. (2025). Mendorong pertumbuhan ekonomi 8% melalui pasar modal Indonesia. Syntax Literate: Jurnal Ilmiah Indonesia, 10(5), 4925–4935. https://doi.org/10.36418/syntax-literate.v10i5.58706

Apridius, W., & Lasut, D. (2025). Implementasi algoritma Random Forest untuk memprediksi harga saham menggunakan indikator Exponential Moving Average 10 periode harian. POTERS: Proceedings of Technology, Engineering and Computers, 1(2). https://jurnal.buddhidharma.ac.id/index.php/poters/article/view/3875

Dewi Lubis, P. K., Br Silalahi, H. H., Sinaga, A. F., Sapma, P. N., & Sitio, V. (2024). Pasar modal dan pengaruhnya terhadap perekonomian di Indonesia. JAKA (Jurnal Akuntansi, Keuangan, dan Auditing), 5(1), 196–214. https://doi.org/10.56696/jaka.v5i1.10755

Ferniawan, M. F., Kusumawati, A., & Madein, A. (2024). The influence of Earnings Per Share (EPS), Price Earnings Ratio (PER), Price to Book Value (PBV), and Debt Equity Ratio (DER) on the stock return. Akrual: Jurnal Bisnis dan Akuntansi Kontemporer, 17(1), 1–14. https://doi.org/10.26487/akrual.v17i01.32073

Mahfuzi, A., & Munawaroh. (2026). Analisa prediksi harga saham Bank BRI berdasarkan data historis menggunakan algoritma Random Forest. Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 6(1), 343–356. https://doi.org/10.55606/teknik.v6i1.10865

Rasyid, A., Muharam, A. B., & Solichin, A. (2025). Prediksi harga saham syariah Indonesia berdasarkan analisis fundamental, teknikal dan bandarmology menggunakan metode Random Forest. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 10(2). https://jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/7855

Sihombing, H. F., Williams, J., Sitorus, J. S., & Sanjaya, M. W. S. (2026). Analisis penerapan nilai Return on Equity (ROE), Price Earning Ratio (PER), Price To Book Value (PBV), dan Debt To Equity Ratio (DER) dalam pengambilan keputusan investasi saham pada sektor perbankan di BEI selama periode 2019–2023. Jurnal Lentera Bisnis, 15(2), 2782–2800.

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 542-552

https://doi.org/10.34127/jrlab.v15i2.2232

Suyatna, S., Satrio, A. J., Pratama, S. P., Indarto, M., Syarif, F., & Amsury, F. (2025). Analisis perbandingan kinerja model machine learning dalam prediksi harga saham TLKM. RIGGS: Journal of Artificial Intelligence and Digital Business, 4(4), 829–837. https://doi.org/10.31004/riggs.v4i4.3476

Sathyanarayanan, S. (2024). Confusion matrix-based performance evaluation metrics. African Journal of Biomedical Research, 27(4), 4023–4031. https://doi.org/10.53555/ajbr.v27i4s.4345

Teixeira, D. M., & Barbosa, R. S. (2025). Stock price prediction in the financial market using machine learning models. Computation, 13(1), 1–24. https://doi.org/10.3390/computation13010003

Urrochman, M. Y., Asy'ari, H., & Hizham, F. A. (2025). Performance comparison of Random Forest regression, SVR models in stock price prediction. Jurnal Pilar Nusa Mandiri, 21(1), 16–23. https://doi.org/10.33480/pilar.v21i1.6072

Zulfa, Z., & Lubis, I. (2025). Pengaruh Earning Per Share dan Price To Book Value terhadap harga saham pada PT Hero Supermarket Tbk periode 2012–2023. Jurnal Ilmiah Manajemen dan Akuntansi, 2(6), 13–23. https://doi.org/10.69714/05j8zf58

Indian Institute of Management Bangalore. (2025). Exploring predictive prowess of ensemble machine learning models in banking stocks: A technical, fundamental, and macroeconomic analysis. IIMB Management Review, 37(2), 100570. https://doi.org/10.1016/j.iimb.2025.100570