

ANALISIS KLASTER RASIO KEUANGAN PT PEGADAIAN DENGAN METODE *K-MEANS* DAN *RANDOM FOREST*

Kimi Thora Refolino¹, Woro Isti Rahayu²

Fakultas Logistik Teknologi dan Bisnis, Universitas Logistik dan Bisnis Internasional¹
email: 184250012@std.ulbi.ac.id¹

Fakultas Logistik Teknologi dan Bisnis, Universitas Logistik dan Bisnis Internasional²
email: woroisti@ulbi.ac.id²

Abstrak

Analisis rasio keuangan banyak digunakan untuk menilai kinerja perusahaan, khususnya dari sisi leverage dan profitabilitas. Namun, perubahan pola rasio keuangan dalam jangka panjang sering sulit dilihat secara jelas tanpa bantuan analisis berbasis data. Penelitian ini bertujuan untuk mengelompokkan kinerja keuangan PT Pegadaian periode 2010–2024 dengan menggabungkan metode *unsupervised learning* dan *supervised learning*. Analisis dilakukan menggunakan rasio Debt to Asset Ratio (DAR) dan Return on Assets (ROA), yang distandardisasi dengan Z-score, kemudian dikelompokkan menggunakan algoritma K-Means dan diuji kembali menggunakan model Random Forest. Hasil penelitian menunjukkan terbentuknya tiga kelompok kinerja keuangan yang menggambarkan perbedaan kondisi leverage dan profitabilitas antarperiode, dengan nilai *silhouette score* sebesar 0,5976. Model Random Forest menghasilkan tingkat akurasi sebesar 80 persen, meskipun kinerjanya lebih rendah pada kelompok dengan jumlah data yang terbatas. **Penelitian ini menunjukkan bahwa kombinasi DAR dan ROA dapat digunakan tidak hanya untuk menilai kinerja keuangan, tetapi juga untuk mengelompokkan dan memahami pola perubahan kinerja perusahaan dalam jangka panjang dengan bantuan metode pembelajaran mesin.**

Kata Kunci: *K-Means, Random Forest, rasio keuangan, Debt to Asset Ratio, Return on Assets.*

Abstract

Financial ratio analysis is widely used to evaluate company performance, particularly in terms of leverage and profitability. However, long-term patterns in financial ratios are often difficult to identify without data-driven analytical support. This study aims to classify the financial performance of PT Pegadaian from 2010 to 2024 by combining unsupervised and supervised learning methods. The analysis uses the Debt to Asset Ratio (DAR) and Return on Assets (ROA), which are standardized using Z-score, clustered using the K-Means algorithm, and validated with a Random Forest model. The results show three distinct financial performance clusters that reflect different leverage and profitability conditions across periods, with a silhouette score of 0.5976. The Random Forest model achieves an accuracy of 80 percent, although its performance is lower for clusters with limited data. **This study shows that DAR and ROA can be used not only to evaluate financial performance, but also to group and identify long-term financial patterns of a company using machine learning methods.**

Keywords: *K-Means, Random Forest, financial ratios, Debt to Asset Ratio, Return on Assets.*

1. PENDAHULUAN

Dalam beberapa tahun terakhir, penggunaan pendekatan ilmu data semakin berkembang dalam analisis keuangan berkat kemampuannya dalam mengolah data historis dan mengidentifikasi pola yang sulit ditangkap oleh metode analisis tradisional. Rasio keuangan seperti Return on Assets (ROA) dan Debt to Asset Ratio (DAR) banyak digunakan untuk mengevaluasi efisiensi aset dan struktur pendanaan perusahaan. Meski penelitian ini mengambil PT Pegadaian sebagai contoh empiris, fokus utama penelitian diarahkan pada penerapan teknik analitik berbasis *machine learning* untuk memahami perubahan jangka panjang rasio keuangan secara lebih sistematis.

Metode K-Means telah banyak digunakan dalam riset keuangan di Indonesia untuk mengelompokkan entitas berdasarkan kemiripan rasio numerik. Wisna, Rani, dan Kastaman menerapkan K-Means untuk menganalisis rasio aktivitas perusahaan subkontraktor dan berhasil mengidentifikasi kelompok perusahaan berdasarkan pola keuangan (Wisna et al., 2023). Penelitian lain oleh Hafidzah menunjukkan bahwa algoritma ini efektif dalam mengelompokkan bank nasional berdasarkan indikator kinerja finansial dan menghasilkan klaster yang berguna untuk penilaian risiko serta pemetaan daya saing institusi keuangan (Hafidzah et al., 2025).

Di sisi lain, algoritma Random Forest banyak digunakan dalam penelitian keuangan untuk membangun model prediksi yang mampu menangani hubungan nonlinier antarvariabel. Hazizah dan rekan-rekan menggunakan Random Forest untuk mengklasifikasikan risiko gagal bayar kartu kredit dan menunjukkan kinerja prediksi yang baik (Hazizah et al., 2025). Meskipun demikian, pemanfaatan Random Forest sebagai alat untuk memverifikasi dan memperkuat hasil klasterisasi rasio keuangan masih jarang dibahas dalam penelitian sebelumnya.

Berdasarkan kondisi tersebut, penelitian ini diperlukan untuk mengisi celah penelitian dengan mengintegrasikan metode *unsupervised learning* dan *supervised learning* dalam analisis rasio keuangan. Penelitian ini menunjukkan pendekatan alternatif dalam memahami dinamika jangka panjang kinerja keuangan perusahaan, di mana K-Means digunakan untuk membentuk segmentasi rasio keuangan, dan Random Forest dimanfaatkan untuk memverifikasi pola klaster yang terbentuk. Dengan demikian, penelitian ini tidak hanya menilai kinerja keuangan secara deskriptif, tetapi juga memberikan kerangka analisis berbasis data yang lebih kuat dan terukur.

2. METODE PENELITIAN

Penelitian ini menerapkan metode kuantitatif dengan penekanan pada pengukuran numerik melalui rasio keuangan Debt to Asset Ratio (DAR) dan Return on Assets (ROA). Seluruh variabel diukur berdasarkan data laporan keuangan dan dianalisis menggunakan teknik *machine learning*. Algoritma K-Means Clustering digunakan untuk mengelompokkan pola kinerja keuangan perusahaan, sedangkan Random Forest digunakan untuk memodelkan dan mengklasifikasikan hasil klaster yang diperoleh dari K-Means. Pendekatan ini sejalan dengan karakteristik metodologi kuantitatif yang mengandalkan data numerik dan analisis statistik terukur (Sugiyono, 2013).

K-Means digunakan untuk mengelompokkan rasio profitabilitas pada perusahaan, termasuk ROA, sebagaimana mereka mengklasifikasikan emiten berdasarkan rasio keuangan (Harianto Kristanto et al., 2016). Sementara itu, *Random Forest* telah digunakan dalam konteks keuangan untuk prediksi risiko dan performa keuangan (Deep et al., 2025). Dalam penelitian ini, Random Forest tidak dimaksudkan untuk memprediksi kinerja finansial secara langsung, melainkan digunakan sebagai alat validasi terhadap hasil klasterisasi yang dihasilkan oleh K-Means.

2.1 Variabel

Analisis kinerja keuangan dalam penelitian ini diawali dengan penghitungan dua rasio utama, yaitu Debt to Asset Ratio (DAR) dan Return on Assets (ROA). Rasio DAR dipilih karena mampu menggambarkan proporsi aset yang dibiayai oleh kewajiban, sehingga menjadi indikator penting dalam menilai tingkat leverage dan risiko struktur modal perusahaan. Penggunaan DAR sebagai instrumen evaluasi struktur pendanaan telah dibahas secara luas dalam literatur manajemen keuangan modern (Pakaya et al., 2023). Secara matematis, DAR dirumuskan sebagai berikut:

$$DAR = \frac{\text{Total Kewajiban}}{\text{Total Asets}} \times 100\%$$

ROA dihitung sebagai perbandingan antara laba bersih dan total aset, yang mencerminkan tingkat efisiensi perusahaan dalam menghasilkan laba dari aset yang dimiliki. Secara sistematis, ROA dapat dirumuskan sebagai berikut:

$$ROA = \frac{\text{Laba Bersih}}{\text{Total Asets}} \times 100\%$$

2.2 Standardisasi Data

Setelah nilai DAR dan ROA dihitung, data kemudian distandardisasi menggunakan metode Z-score untuk menyamakan skala antarvariabel sebelum diproses oleh algoritma *machine learning*. Standardisasi ini menjadi tahap yang krusial mengingat perbedaan skala dapat memengaruhi sensitivitas pengukuran jarak pada algoritma K-Means serta proses pemisahan fitur pada Random Forest. Dengan demikian, normalisasi berperan dalam meningkatkan stabilitas dan akurasi model yang digunakan (Johor Bahru et al., 2013). Formula Z-score dirumuskan sebagai:

$$Z = \frac{X - \mu}{\sigma}$$

di mana X adalah nilai rasio suatu tahun, μ adalah rata-rata rasio sepanjang periode, dan σ adalah simpangan baku.

2.3 K-Means Cluster

Pembentukan kluster dilakukan menggunakan algoritma K-Means, yaitu metode *unsupervised learning* yang bertujuan mengelompokkan sekumpulan observasi berdasarkan tingkat kemiripan karakteristik numeriknya. Algoritma ini dipilih karena memiliki efisiensi komputasi yang tinggi pada data berskala kontinu serta menunjukkan performa yang stabil dalam berbagai penelitian keuangan dan analisis data sains, termasuk dalam pengelompokan rasio keuangan dan evaluasi kinerja perusahaan (Aditya et al., 2024; Mulyaningsih & Heikal, 2022).

K-Means beroperasi dengan meminimalkan jarak antara setiap titik data dan pusat kluster (*centroid*) terdekat. Proses ini berlangsung secara iteratif, dimulai dari inisialisasi centroid awal, pengelompokan data berdasarkan jarak Euclidean, hingga mencapai kondisi konvergen ketika perubahan posisi centroid tidak lagi signifikan (Aditya et al., 2024). Dalam penelitian ini, fitur yang digunakan untuk pembentukan kluster adalah nilai hasil standardisasi DAR dan ROA menggunakan Z-score, sehingga setiap titik data merepresentasikan profil *leverage* dan profitabilitas suatu perusahaan pada periode tertentu. Rumus utama K-Means melibatkan perhitungan jarak *Euclidean* sebagai ukuran kedekatan antara data dan centroid, dituliskan sebagai:

$$d(x_i, C_j) = \sqrt{\sum_{k=1}^n (x_{ik} - C_{jk})^2}$$

di mana x_i adalah data ke- i , C_j adalah centroid kluster ke- j , dan n adalah jumlah variabel fitur yang dimana dalam penelitian ini adalah DAR dan ROA. Selanjutnya, proses optimasi dilakukan dengan

meminimalkan fungsi objektif *Within-Cluster Sum of Squares* (WCSS), yang dirumuskan sebagai:

$$WCSS = \sum_{j=1}^K \sum_{x_i \in C_j} \|x_i - C_j\|^2$$

Semakin kecil nilai WCSS, semakin baik kualitas kluster karena setiap data berada dekat dengan pusat klusternya. *K-Means* akan terus memperbarui centroid menggunakan rata-rata data dalam kluster. Semakin kecil nilai WCSS, semakin baik kualitas kluster karena setiap data berada dekat dengan pusat klusternya. *K-Means* akan terus memperbarui centroid menggunakan rata-rata data dalam kluster dan proses ini terus berulang hingga konvergen.

2.4 *Random Forest*

Pembentukan model *Random Forest* pada penelitian ini dilakukan sebagai metode *supervised learning* yang mampu menghasilkan prediksi lebih stabil melalui proses agregasi sejumlah pohon keputusan. Algoritma ini bekerja dengan membangun banyak decision tree menggunakan subset data dan subset fitur yang dipilih secara acak (*bootstrap aggregation*), kemudian menentukan keluaran akhir berdasarkan mekanisme majority voting untuk klasifikasi atau averaging untuk regresi. Proses bagging ini bertujuan mengurangi varians model sehingga risiko *overfitting* menjadi lebih rendah dibandingkan satu pohon keputusan tunggal (Jin, 2025). Secara matematis, prediksi akhir *Random Forest* untuk kasus regresi dihitung sebagai rata-rata keluaran seluruh pohon, yang dirumuskan sebagai:

$$\hat{y}^{RF} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t$$

sedangkan untuk klasifikasi, label akhir ditentukan oleh kelas yang paling sering muncul dari seluruh pohon, yaitu:

$$\hat{y} = \text{mode} \{f_1(x), f_2(x), \dots, f_n(x)\}$$

Dalam penelitian ini, model *Random Forest* digunakan untuk mengevaluasi kontribusi rasio keuangan, khususnya DAR dan ROA, terhadap proses pengelompokan yang sebelumnya dihasilkan oleh *K-Means*, sekaligus menilai stabilitas pemisahan data berdasarkan karakteristik *leverage* dan profitabilitas perusahaan. Pendekatan ini memberikan gambaran komprehensif mengenai pola klasifikasi finansial serta memperkuat validitas analisis melalui kombinasi metode *unsupervised* dan *supervised learning* (Li, 2025; Liu et al., 2015).

2.5 Konfigurasi Model dan Data set

Dalam tahap implementasi model, dataset yang terdiri dari 15 observasi tahunan dibagi menjadi data latih dan data uji dengan rasio 70:30, di mana 70 persen data digunakan untuk melatih model dan 30 persen sisanya digunakan untuk menguji performa klasifikasi. Pembagian data dilakukan secara terstratifikasi untuk menjaga proporsi kluster tetap seimbang antara data latih dan data uji. Model *Random Forest* dikonfigurasi menggunakan 200 pohon keputusan ($n_estimators = 200$) dengan kedalaman maksimum pohon sebesar 5 ($max_depth = 5$). Pengaturan parameter ini bertujuan untuk memperoleh keseimbangan antara kemampuan generalisasi model dan kompleksitas struktur pohon, sehingga risiko *overfitting* dapat ditekan, khususnya pada dataset dengan ukuran terbatas. Evaluasi performa model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, serta diperkuat dengan validasi silang tiga lipatan (*3-fold cross-validation*) untuk menilai konsistensi kinerja model.

Seluruh proses analisis data, mulai dari perhitungan rasio keuangan, standarisasi data, pembentukan kluster *K-Means*, hingga evaluasi model *Random Forest*, diimplementasikan menggunakan bahasa pemrograman Python dengan dukungan pustaka *pandas*, *NumPy*, *scikit-learn*, dan *matplotlib*.

2.6 Data Limitation

Penelitian ini menggunakan data keuangan tahunan yang terdiri dari total aset, total kewajiban, dan laba bersih PT Pegadaian selama periode 15 tahun. Jumlah observasi yang relatif terbatas ini berimplikasi langsung terhadap performa dan stabilitas metode *supervised learning* yang digunakan, khususnya Random Forest. Pada ukuran data yang kecil, model cenderung memiliki keterbatasan dalam menangkap variasi pola yang kompleks, sehingga hasil klasifikasi lebih merepresentasikan kecenderungan umum dibandingkan pola yang sangat rinci.

Keterbatasan jumlah data juga berpotensi meningkatkan sensitivitas model terhadap pembagian data latih dan data uji, meskipun dalam penelitian ini telah diterapkan pembagian data secara terstratifikasi serta validasi silang untuk meminimalkan bias. Oleh karena itu, penggunaan Random Forest dalam penelitian ini tidak dimaksudkan untuk menghasilkan prediksi finansial jangka panjang, melainkan sebagai alat validasi kluster hasil K-Means serta untuk mengevaluasi kontribusi rasio keuangan DAR dan ROA terhadap pembentukan kelompok kinerja keuangan.

Selain itu, penggunaan data agregat tahunan menyebabkan dinamika keuangan jangka pendek tidak dapat diamati secara detail. Dengan demikian, hasil penelitian ini perlu diinterpretasikan secara hati-hati dan lebih tepat dipahami sebagai analisis eksploratif terhadap pola leverage dan profitabilitas perusahaan. Penelitian selanjutnya disarankan untuk menggunakan data dengan frekuensi yang lebih tinggi atau jumlah observasi yang lebih besar guna meningkatkan robustitas model *supervised learning*.

3. HASIL DAN PEMBAHASAN

Analisis hasil penelitian memanfaatkan empat variabel utama yaitu aset, kewajiban, dan laba bersih yang terlebih dahulu disusun dalam bentuk numerik untuk memastikan ketelitian pada tahap pengolahan berikutnya. Variabel-variabel tersebut menjadi dasar perhitungan dua rasio keuangan, yaitu *Debt to Asset Ratio* (DAR) sebagai indikator struktur pendanaan dan *Return on Assets* (ROA) sebagai ukuran efektivitas pemanfaatan aset dalam menghasilkan laba. Berikut data keuangan PT Pegadaian.

Tabel 1. Data Keuangan PT Pegadaian Tahun 2010-2024

Tahun	Aset	Kewajiban	Laba Bersih
2010	Rp 20.283.043	Rp 16.986.840	Rp 1.179.788
2011	Rp 26.219.352	Rp 22.142.989	Rp 1.467.325
2012	Rp 29.311.898	Rp 23.929.784	Rp 1.904.822
2013	Rp 33.469.357	Rp 23.110.735	Rp 1.907.304
2014	Rp 35.443.361	Rp 24.634.255	Rp 1.761.764
2015	Rp 39.157.960	Rp 26.625.016	Rp 1.938.429
2016	Rp 46.873.892	Rp 30.396.979	Rp 2.210.252
2017	Rp 48.687.092	Rp 30.476.832	Rp 2.513.539
2018	Rp 52.791.188	Rp 32.674.699	Rp 2.775.481
2019	Rp 65.324.177	Rp 42.263.867	Rp 3.108.078
2020	Rp 71.468.960	Rp 46.865.344	Rp 2.022.447
2021	Rp 65.775.938	Rp 39.516.937	Rp 2.427.310
2022	Rp 73.329.790	Rp 44.911.304	Rp 3.298.945
2023	Rp 82.585.431	Rp 49.949.840	Rp 4.376.677
2024	Rp 102.616.631	Rp 66.642.222	Rp 5.851.797

3.1 Penghitungan DAR dan ROA

Rasio DAR diperoleh dari perbandingan total kewajiban terhadap total aset, sedangkan ROA dihitung melalui pembagian laba bersih terhadap total aset. Seluruh nilai kemudian dikonversi ke dalam persentase untuk memudahkan interpretasi dan komparabilitas antarperiode. Kedua rasio dihitung menggunakan data tahun 2010–2024 melalui pemrograman Python, sehingga menghasilkan ukuran yang lebih konsisten.

```
df['DAR'] = df['Kewajiban'] / df['Aset'] * 100  
df['ROA'] = df['Laba Bersih'] / df['Aset'] * 100  
print(df[['Tahun', 'DAR', 'ROA']])
```

Gambar 1. Kode DAR dan ROA

3.2 Standardisasi Data

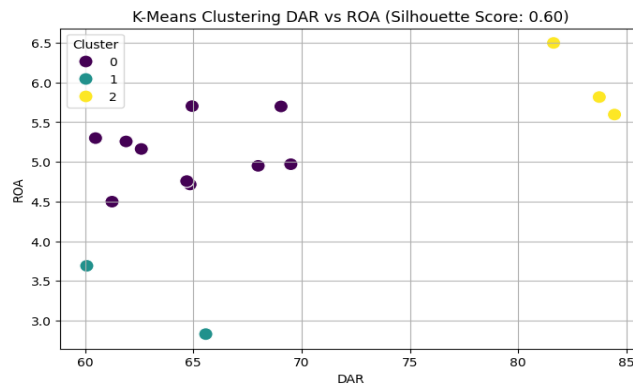
Standardisasi data dilakukan terhadap rasio Debt to Asset Ratio (DAR) dan Return on Assets (ROA) menggunakan metode *Z-score* untuk menyamakan skala kedua variabel sebelum proses klusterisasi. Proses ini menghasilkan variabel *DAR_Zscore* dan *ROA_Zscore* yang merepresentasikan posisi relatif masing-masing tahun terhadap rata-rata periode pengamatan 2010–2024. Standardisasi dilakukan menggunakan pemrograman Python sehingga setiap rasio memiliki kontribusi yang seimbang dalam perhitungan jarak pada algoritma *K-Means* serta mendukung stabilitas analisis pada tahap selanjutnya.

Tabel 2. Nilai Rasio Keuangan dan Hasil Standardisasi *Z-score* PT Pegadaian Periode 2010–2024

Tahun	DAR (%)	ROA (%)	DAR_Zscore	ROA_Zscore
2010	83.7490	5.8166	1.9307	0.9091
2011	84.4528	5.5963	2.0180	0.6547
2012	81.6385	6.4985	1.6689	1.6969
2013	69.0504	5.6987	0.1075	0.7729
2014	69.5032	4.9706	0.1637	-0.0682
2015	67.9939	4.9503	-0.0235	-0.0917
2016	64.8484	4.7153	-0.4136	-0.3632
2017	62.5974	5.1626	-0.6929	0.1536
2018	61.8942	5.2575	-0.7801	0.2632
2019	64.6987	4.7579	-0.4322	-0.3140
2020	65.5744	2.8298	-0.3236	-2.5415
2021	60.0781	3.6903	-1.0053	-1.5474
2022	61.2456	4.4988	-0.8605	-0.6134
2023	60.4826	5.2996	-0.9552	0.3118
2024	64.9429	5.7026	-0.4019	0.7774

3.3 *K-Means Cluster*

Algoritma *K-Means* diterapkan untuk mengelompokkan data rasio keuangan berdasarkan kemiripan nilai DAR dan ROA, dengan terlebih dahulu melakukan normalisasi menggunakan *StandardScaler* agar kedua variabel berada pada skala yang sebanding dan tidak menimbulkan bias dalam pembentukan pusat kluster. Setelah proses standardisasi, model ditetapkan untuk menghasilkan tiga kluster (*n_Clusters* = 3) sesuai kebutuhan analisis rasio keuangan, kemudian dilakukan proses *fitting* sehingga setiap observasi memperoleh label kluster berdasarkan kedekatannya dengan centroid yang terbentuk..



Gambar 2. Hasil dan grafik dari Cluster

Hasil klasterisasi K-Means menunjukkan bahwa data keuangan perusahaan terbagi ke dalam tiga kelompok dengan karakteristik Debt to Asset Ratio (DAR) dan Return on Assets (ROA) yang berbeda, sehingga pola kinerja keuangan antarperiode dapat diamati secara lebih jelas. Cluster 0 mencakup sebagian besar tahun 2013–2024 (kecuali 2020–2021) dan menggambarkan kondisi keuangan yang relatif stabil, dengan DAR berada pada tingkat menengah sekitar 60–70 persen serta ROA yang berada pada level menengah dan cenderung konsisten. Pola ini mencerminkan periode stabilisasi struktur modal dan profitabilitas yang masih terjaga.

Cluster 2 yang terdiri dari tahun 2010–2012 menunjukkan tingkat leverage yang sangat tinggi dengan DAR mencapai kisaran 82–84 persen. Meskipun struktur modal berada pada tekanan yang cukup besar, ROA pada periode ini tetap relatif tinggi dan stabil, yang menandakan bahwa perusahaan masih mampu menghasilkan laba secara efisien. Sebaliknya, Cluster 1 yang mencakup tahun 2020–2021 memperlihatkan kondisi yang paling tertekan, dengan DAR berada pada tingkat menengah namun ROA mengalami penurunan tajam, khususnya pada tahun 2020, yang mengindikasikan tekanan profitabilitas akibat faktor eksternal.

Pemisahan klaster tersebut dapat terbentuk dengan jelas karena kombinasi DAR dan ROA merepresentasikan dua aspek utama kondisi keuangan perusahaan, yaitu struktur pendanaan dan kemampuan menghasilkan laba dari aset. DAR menggambarkan proporsi aset yang dibiayai oleh kewajiban, sedangkan ROA menunjukkan tingkat efisiensi aset dalam menghasilkan laba. Perbedaan nilai kedua rasio ini antarperiode menciptakan pola yang kontras, sehingga algoritma K-Means mampu mengelompokkan tahun-tahun dengan karakteristik keuangan yang serupa secara konsisten. Hal ini didukung oleh nilai silhouette score sebesar 0,5976 yang menunjukkan bahwa batas antarcluster cukup tegas, serta hasil Random Forest yang mampu mengenali klaster dengan tingkat akurasi yang memadai. Temuan ini menegaskan bahwa variasi leverage dan profitabilitas merupakan faktor utama dalam pembentukan segmentasi kinerja keuangan perusahaan.

3.4 Random Forest

Penerapan algoritma *Random Forest* dalam penelitian ini bertujuan untuk mengevaluasi kemampuan model klasifikasi dalam mengidentifikasi klaster hasil pengelompokan *K-Means* berdasarkan dua variabel utama, yaitu *Debt to Asset Ratio* (DAR) dan *Return on Assets* (ROA). Pendekatan ini memberikan gambaran mengenai sejauh mana struktur pola klaster yang terbentuk dapat dikenali kembali oleh metode klasifikasi berbasis ensemble, sehingga keandalan segmentasi keuangan yang diperoleh dapat diuji secara lebih baik.

```
rf = RandomForestClassifier(  
    n_estimators=200,  
    max_depth=5,  
    random_state=42  
)  
  
rf.fit(X_train, y_train)  
y_pred = rf.predict(X_test)
```

Gambar 3. Kode algoritma *Random Forest*

Hasil evaluasi menggunakan classification report menunjukkan bahwa model *Random Forest* mencapai tingkat akurasi keseluruhan sebesar **80%** pada data uji. Metrik precision, recall, dan *f1-score* memperlihatkan variasi kinerja antar-klaster, yang berkaitan erat dengan distribusi jumlah data pada masing-masing kelompok.

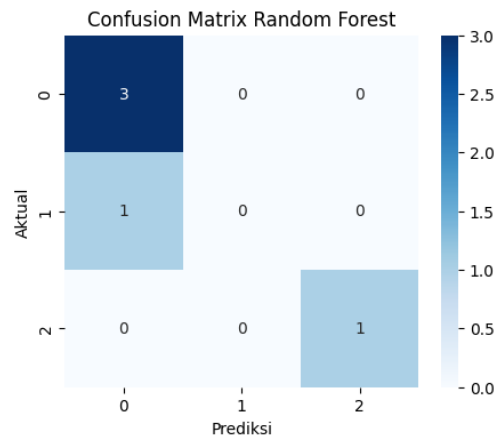
Tabel 3. Hasil metric *Random Forest*

Metrik	Value
Accuracy	0.8000
Precision	0.9167
Recall	0.6667
F-1 Score	0.6190

Model *Random Forest* yang digunakan untuk memvalidasi hasil klaster *K-Means* menghasilkan nilai akurasi sebesar **0,80**, yang menunjukkan bahwa **4 dari 5 observasi data uji** dapat diklasifikasikan ke dalam klaster yang sesuai. Nilai **precision sebesar 0,9167** mengindikasikan bahwa lebih dari **91% prediksi klaster yang dihasilkan model bersifat tepat**, sehingga kesalahan klasifikasi relatif rendah pada klaster yang berhasil dikenali. Namun demikian, nilai **recall sebesar 0,6667** menunjukkan bahwa hanya sekitar **dua pertiga anggota klaster aktual** yang dapat diidentifikasi dengan benar oleh model, yang berdampak pada nilai **F1-score sebesar 0,6190** sebagai ukuran keseimbangan antara precision dan recall. Kondisi ini mencerminkan keterbatasan jumlah data uji yang relatif kecil serta distribusi observasi antar klaster yang tidak sepenuhnya seimbang.

Proses validasi silang menggunakan skema 3-fold menghasilkan skor rata-rata sebesar 0.8000. Nilai ini mengindikasikan bahwa model memiliki tingkat kestabilan yang baik ketika diuji pada berbagai subset data. Stabilitas ini memperkuat asumsi bahwa struktur klaster yang dibentuk memiliki pola yang relatif jelas terhadap variabel DAR dan ROA.

Analisis *confusion matrix* memberikan gambaran lebih rinci mengenai distribusi kesalahan prediksi. Untuk klaster 0 dan 2, model menunjukkan tingkat prediksi yang akurat tanpa kesalahan klasifikasi. Namun, klaster 1 sepenuhnya diprediksi sebagai kelas lain sehingga tidak ada data yang teridentifikasi benar. Temuan ini mengonfirmasi bahwa model mengalami bias terhadap kelas dominan, akibat perbedaan proporsi data yang terlalu besar antar-klaster.



Gambar 4. Confusion Matrix

Confusion matrix Random Forest menunjukkan bahwa dari total data uji, terdapat **3 observasi pada klaster pertama** dan **1 observasi pada klaster ketiga** yang berhasil diklasifikasikan dengan benar, sementara **1 observasi pada klaster kedua** mengalami kesalahan klasifikasi. Pola ini terlihat dari nilai diagonal utama matriks yang menunjukkan prediksi benar, yaitu **3 pada klaster 0** dan **1 pada klaster 2**, serta nilai nol pada sebagian besar elemen non-diagonal, yang menandakan rendahnya tingkat kesalahan lintas klaster.

Kesalahan klasifikasi yang terjadi pada klaster kedua mengindikasikan bahwa model mengalami keterbatasan dalam mengenali karakteristik klaster dengan jumlah sampel yang sangat kecil pada data uji. Kondisi ini memperkuat temuan bahwa Random Forest dalam penelitian ini lebih tepat digunakan sebagai alat validasi hasil klaster K-Means daripada sebagai model klasifikasi yang berdiri sendiri, terutama mengingat ukuran data yang terbatas

4. KESIMPULAN DAN SARAN

Hasil analisis menunjukkan bahwa kombinasi rasio Debt to Asset Ratio (DAR) dan Return on Assets (ROA) mampu membentuk segmentasi kinerja keuangan perusahaan secara jelas dan konsisten, ditandai oleh terbentuknya tiga klaster dengan karakteristik leverage dan profitabilitas yang berbeda. Segmentasi ini memberikan gambaran ringkas namun informatif mengenai kondisi keuangan antarperiode, sehingga memudahkan interpretasi pola stabilitas, tekanan, dan efisiensi kinerja. Pendekatan yang digunakan bersifat sederhana, terukur, dan mudah direplikasi, sehingga berpotensi diterapkan secara berkelanjutan untuk pemantauan kinerja keuangan pada periode berikutnya maupun pada entitas dengan karakteristik serupa.

5. REFERENSI

- Aditya, G., Firdaus, A., & Wulandhari, L. A. (2024). Company clustering based on financial report data using k-means. *Computer Science and Information Technologies*, 5(3), 243–253. <https://doi.org/10.11591/csit.v5i3.pp243-253>
- Deep, A., Shirvani, A., Monico, C., Rachev, S., & Fabozzi, F. (2025). Risk-Adjusted Performance of Random Forest Models in High-Frequency Trading. *Journal of Risk and Financial Management*, 18(3). <https://doi.org/10.3390/jrfm18030142>
- Hafidzah, A. N., Gulo, H. S., Yolanda, S. P., Saban, V. P. A., Sari, V. A., & Nanang, N. (2025). Klasterisasi Bank di Indonesia Berdasarkan Kinerja Finansial Menggunakan Algoritma K-Means. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 5075–5080. <https://doi.org/10.31004/riggs.v4i2.1387>

- Harianto Kristanto, N., Christopher, A. L., & Budi, H. S. (2016). Implementasi K-Means Clustering untuk Pengelompokan Analisis Rasio Profitabilitas dalam Working Capital. *JUI SI*, 02(01).
- Hazizah, N., Sharipudin, & Feranika, A. (2025). Implementasi Algoritma Random Forest Dalam Klasifikasi Risiko Gagal Bayar Kartu Kredit Pada Nasabah Bank. *Jurnal Manajemen Teknologi Dan Sistem Informasi (JMS)*, 5(1). <https://doi.org/10.33998/jms.v5i1>
- Jin, Y. (2025). *Unlocking Stock Return Predictions: Using Financial Statements with Random Forest and PCA* (pp. 664–673). https://doi.org/10.2991/978-94-6463-652-9_69
- Johor Bahru, U., Darul Ta, J., Bin Mohamad, I., Usman, D., & Bahru, J. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), 3299–3303.
- Li, Y. (2025). *Optimizing Random Forest Models for Stock Market Prediction with Hyperparameter Analysis*. 57–61. <https://doi.org/10.1145/3745238.3745250>
- Liu, C., Chan, Y., Alam Kazmi, S. H., & Fu, H. (2015). Financial Fraud Detection Model: Based on Random Forest. *International Journal of Economics and Finance*, 7(7). <https://doi.org/10.5539/ijef.v7n7p178>
- Mulyaningsih, S., & Heikal, J. (2022). K-Means Clustering Using Principal Component Analysis (PCA) Indonesia Multi-Finance Industry Performance Before and During Covid-19. *Asia Pacific Management and Business Application*, 011(02), 131–142. <https://doi.org/10.21776/ub.apmba.2022.011.02.1>
- Pakaya, A., Lamuda, I., & Rahman, Z. (2023). ANALISIS RASIO LEVERAGE UNTUK MENILAI KINERJA KEUANGAN PERUSAHAAN DAGANG. *JEMAI: Jurnal Ekonomi Manajemen Dan Akuntansi*, 2(2). <https://jurnal.unigo.ac.id/index.php/jemai>
- Sugiyono. (2013). *METODE PENELITIAN KUANTITATIF KUALITATIF DAN R&D*.
- Wisna, N., Cahaya Rani, M., & Kastaman. (2023). ALGORITMA K-MEANS CLUSTERING ANALISIS RASIO AKTIVITAS MENGGUNAKAN PYTHON. *Ecobisma: Jurnal Ekonomi, Bisnis, Dan Manajemen*, 10(2). <https://doi.org/10.36987/ecobi.v10i2>