Analisis Dampak Penerapan SMOTE terhadap Performa Prediksi dan Stabilitas *Feature Importance Random Forest* pada *Financial Distress*

Analyzing the Impact of SMOTE on Predictive Performance and Random Forest Feature Importance Stability in Financial Distress Prediction

[1]Alan Kusuma Soesanto*, [2]Charitas Fibriani
[1,2]Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
[1,2]Jl. Dr. O. Notohamidjojo No. 1-10, Blotongan, Sidorejo, Salatiga, Jawa Tengah 501715 Indonesia
*e-mail: charitas.fibriani@uksw.edu

(***received*:** 10 September 2026, ***revised*:** 15 October 2026, ***accepted*:** 16 October 2026)

Abstrak

Prediksi *financial distress* sering terkendala ketidakseimbangan kelas yang membuat model klasifikasi bias terhadap kelas mayoritas. *Synthetic Minority Oversampling Technique* (SMOTE) banyak digunakan untuk meningkatkan deteksi kelas minoritas, tetapi dampaknya terhadap konsistensi *feature importance* sebagai dasar interpretasi model masih jarang dikaji. Penelitian ini menganalisis dampak penerapan SMOTE terhadap performa prediksi dan stabilitas *feature importance Random Forest* menggunakan dataset kebangkrutan perusahaan Taiwan (6.819 perusahaan, 220 di antaranya bangkrut; 39 rasio keuangan setelah seleksi fitur). Tiga skenario dibandingkan, yaitu *Random Forest* tanpa penanganan ketidakseimbangan (RF), dengan pembobotan kelas (RF + CW), dan dengan SMOTE yang diterapkan di dalam *pipeline cross-validation* (RF + SMOTE). Ketiga skenario dioptimasi dengan ruang pencarian *hyperparameter* yang sama melalui GridSearchCV 5-fold dan metrik F2-Score. Perbedaan performa pada data uji diuji dengan *paired bootstrap* dan uji McNemar, sedangkan stabilitas *feature importance* diukur dengan korelasi Spearman, Kendall's Tau, dan kesamaan Jaccard lintas 10 *random seed*. Dibandingkan RF, SMOTE meningkatkan *Recall* secara signifikan dari 0,2045 menjadi 0,6818 (CI 95% selisih: 0,341 s.d. 0,636), tetapi menurunkan *Precision* dari 0,8182 menjadi 0,3659; perubahan ROC-AUC (0,9439 menjadi 0,9551) dan PR-AUC tidak signifikan. Performa RF + SMOTE juga tidak berbeda signifikan dengan RF + CW. Peringkat *feature importance* berubah secara moderat setelah SMOTE (Spearman 0,644 ± 0,017), lebih besar daripada variasi antar-seed (0,970), namun delapan rasio, antara lain *Total debt/Total net worth* dan *Persistent EPS in the Last Four Seasons*, konsisten berada pada Top-10 di seluruh pengulangan. Hasil ini menunjukkan bahwa SMOTE terutama menggeser batas keputusan model ke arah deteksi kelas minoritas, bukan meningkatkan kemampuan diskriminasi, sehingga interpretasi *feature importance* sebaiknya difokuskan pada rasio yang stabil lintas skenario.

Kata kunci: feature importance, financial distress, stabilitas data tidak seimbang, Random Forest, SMOTE.

Abstract

Financial distress prediction is often challenged by class imbalance, which can cause classification models to become biased toward the majority class. The Synthetic Minority Over-sampling Technique (SMOTE) is widely used to improve minority-class detection; however, its impact on the consistency of feature importance as a basis for model interpretation remains relatively underexplored. This study examines the impact of SMOTE on predictive performance and feature importance stability in Random Forest models using the Taiwan corporate bankruptcy dataset, comprising 6,819 companies, of which 220 were bankrupt, and 39 financial ratios retained after feature selection. Three scenarios were compared: Random Forest without class imbalance treatment (RF), Random Forest with class weighting (RF + CW), and Random Forest with SMOTE applied within the cross-validation pipeline (RF + SMOTE). All three scenarios were optimized using the same hyperparameter search space through five-fold GridSearchCV, with the F2-score as the optimization metric. Differences in test-set

 ISSN:2302-8149
e-ISSN:2540-9719

performance were evaluated using paired bootstrap and McNemar's test, while feature importance stability was assessed using Spearman's rank correlation, Kendall's tau, and Jaccard similarity across 10 random seeds. Compared with RF, SMOTE significantly increased recall from 0.2045 to 0.6818 (95% CI for the difference: 0.341 to 0.636), but reduced precision from 0.8182 to 0.3659. Changes in ROC-AUC (from 0.9439 to 0.9551) and PR-AUC were not statistically significant. The predictive performance of RF + SMOTE also did not differ significantly from that of RF + CW. Feature importance rankings changed moderately after SMOTE was applied (Spearman's correlation: 0.644 ± 0.017), indicating greater variation than that observed across random seeds (0.970). Nevertheless, eight financial ratios, including Total Debt/Total Net Worth and Persistent EPS in the Last Four Seasons, consistently appeared among the top 10 features across all repetitions. These findings indicate that SMOTE primarily shifts the model's decision boundary toward greater minority-class detection rather than improving its discriminatory ability. Therefore, the interpretation of feature importance should prioritize financial ratios that remain stable across different modeling scenarios.
Keywords: *feature importance stability, financial distress, imbalanced data, Random Forest, SMOTE.*

1 Pendahuluan

Financial distress merupakan masalah penting dalam dunia bisnis karena berdampak luas tidak hanya bagi manajemen internal perusahaan, tetapi juga bagi investor, kreditor, dan pemangku kepentingan lainnya. Kegagalan mendeteksi tanda-tanda *distress* sejak dini dapat mengakibatkan kerugian finansial yang besar, sehingga model prediksi yang andal diperlukan sebagai sistem peringatan dini (*early warning system*) dan pendukung pengambilan keputusan bisnis [12]. Algoritma *Random Forest* banyak digunakan untuk kebutuhan ini karena performanya yang baik pada data keuangan [1], [7], [10], [11]. Penelitian [1] menunjukkan *Random Forest* menghasilkan F1-Score dan AUC lebih tinggi dibandingkan *Logistic Regression* pada sistem *credit scoring*, sedangkan [25] menemukan bahwa model berbasis *ensemble* secara konsisten mengungguli model statistik konvensional dalam prediksi kebangkrutan.

Dataset *financial distress* pada umumnya bersifat *highly imbalanced* karena jumlah perusahaan sehat jauh lebih banyak dibandingkan perusahaan yang mengalami *distress*. Kondisi ini menyebabkan pengklasifikasi cenderung bias terhadap kelas mayoritas sehingga kemampuan mendeteksi perusahaan *distress* (*Recall*) menjadi rendah [2], [3]. SMOTE [16] merupakan salah satu teknik penyeimbangan data yang paling banyak digunakan untuk mengatasi masalah tersebut, yaitu dengan membangkitkan sampel sintetis kelas minoritas melalui interpolasi antara sampel minoritas dan tetangga terdekatnya [17]. Sejumlah penelitian melaporkan bahwa SMOTE meningkatkan *Recall* pada prediksi kebangkrutan [2], [14], sementara [3] menunjukkan bahwa SMOTEBoost dan SMOTEBagging menghasilkan *recall* dan akurasi lebih tinggi dibandingkan teknik *undersampling*. Meskipun demikian, [24] mengingatkan bahwa koreksi ketidakseimbangan kelas juga mengubah distribusi probabilitas keluaran model, sehingga dampaknya perlu dievaluasi secara cermat dan tidak hanya dari satu metrik.

Sebagian besar penelitian sebelumnya mengevaluasi SMOTE dari sisi performa prediksi, seperti *Accuracy*, *Recall*, F1-Score, atau ROC-AUC [4], [13], [14]. Di sisi lain, interpretasi model melalui *feature importance* sering dijadikan dasar rekomendasi rasio keuangan. Penelitian [5] mengidentifikasi *debt-to-equity ratio*, *current ratio*, dan *return on assets* sebagai fitur terpenting, dan [6] mengidentifikasi *Return on Assets* sebagai variabel paling dominan pada *Decision Tree*, namun kedua penelitian tersebut mengukur *feature importance* pada satu kondisi data saja. Literatur *machine learning* menunjukkan bahwa *feature importance Random Forest* berbasis *Mean Decrease in Impurity* (MDI) dapat bias [19] dan bahwa peringkat fitur dapat tidak stabil terhadap perubahan data latih [21]. Karena SMOTE mengubah distribusi data latih dengan menambahkan sampel sintetis, peringkat *feature importance* berpotensi ikut berubah. Berdasarkan penelusuran literatur yang dilakukan penulis, kajian yang mengukur secara kuantitatif stabilitas peringkat *feature importance Random Forest* sebelum dan sesudah SMOTE pada konteks prediksi *financial distress* masih terbatas.

Berdasarkan kesenjangan tersebut, rumusan masalah penelitian ini adalah: (1) bagaimana dampak penerapan SMOTE terhadap performa prediksi *Random Forest* pada klasifikasi *financial distress*; dan (2) bagaimana dampak penerapan SMOTE terhadap stabilitas *feature importance*

 ISSN:2302-8149
e-ISSN:2540-9719

Random Forest dalam mengidentifikasi rasio keuangan yang penting. Sejalan dengan rumusan masalah tersebut, tujuan penelitian ini adalah:

1. Menganalisis dampak penerapan SMOTE terhadap performa prediksi *Random Forest* dibandingkan model tanpa penanganan ketidakseimbangan dan model dengan pembobotan kelas, disertai uji signifikansi statistik.
2. Menganalisis stabilitas peringkat *feature importance Random Forest* sebelum dan sesudah penerapan SMOTE, serta membedakan pergeseran peringkat yang disebabkan SMOTE dari variasi acak model.

Kontribusi penelitian ini adalah evaluasi dampak SMOTE secara bersamaan dari aspek performa klasifikasi dan stabilitas interpretasi model dengan desain eksperimen terkontrol, sehingga dapat memberikan panduan pemilihan rasio keuangan yang lebih reliabel bagi manajemen dan investor.

2 Tinjauan Literatur

Random Forest [18] membangun banyak pohon keputusan dari sampel *bootstrap* dan subset fitur acak, lalu menggabungkan hasilnya melalui *voting*, sehingga mampu menangkap hubungan non-linear antar rasio keuangan. Keunggulan ini tercermin pada berbagai studi prediksi kebangkrutan: [2] melaporkan selisih akurasi yang besar antara *Random Forest* dan *Logistic Regression* (95% berbanding 57%), [7] mencapai akurasi 97,8%, dan [10], [13] juga menempatkan *Random Forest* sebagai salah satu model terbaik dibanding algoritma lain. Pada konteks Indonesia, [11] menunjukkan bahwa model *machine learning* mampu memprediksi *financial distress* perusahaan secara akurat. Namun, sebagian besar studi tersebut menggunakan akurasi sebagai metrik utama, yang dapat menyesatkan pada data tidak seimbang.

Ketidakseimbangan kelas merupakan karakteristik khas data kebangkrutan; [3] melaporkan rasio *imbalance* ekstrem hingga 0,02. SMOTE [16] mengatasinya dengan membangkitkan sampel sintetis di sepanjang garis antara sampel minoritas dan *k* tetangga terdekatnya, dan telah berkembang menjadi pendekatan standar penanganan data tidak seimbang [17]. Penerapannya memiliki dua catatan metodologis. Pertama, *oversampling* yang dilakukan sebelum pembentukan *fold cross-validation* menyebabkan informasi dari sampel yang sama muncul pada *training* dan *validation fold*, sehingga estimasi performa menjadi terlalu optimistis (*data leakage*) [20]. Kedua, koreksi ketidakseimbangan kelas menggeser probabilitas prediksi, sehingga perbandingan antar-model perlu memisahkan efek penyeimbangan data dari efek pemilihan *threshold* [24]. Dari sisi evaluasi, [1] mengusulkan F1-Score dan AUC sebagai pelengkap akurasi, [4] menggunakan tujuh metrik sekaligus, dan [3] menegaskan pentingnya *recall* karena kegagalan mendeteksi kasus positif menimbulkan konsekuensi bisnis yang lebih besar.

Terkait interpretasi model, *Random Forest* menyediakan *feature importance* berbasis MDI, yaitu rata-rata penurunan *impurity* yang dihasilkan suatu fitur pada seluruh pohon [18]. Ukuran ini mudah diperoleh, tetapi diketahui bias terhadap fitur kontinu atau fitur dengan banyak nilai unik [19], sehingga *permutation importance*, yaitu penurunan performa ketika nilai suatu fitur diacak, sering digunakan sebagai pembanding [23]. Pada domain keuangan, [5] menemukan *debt-to-equity ratio*, *current ratio*, dan *return on assets* sebagai fitur terpenting, [6] menegaskan dominasi *Return on Assets* pada *Decision Tree*, pendekatan Altman Z-Score dan Springate pada [8] menghasilkan klasifikasi yang berbeda untuk perusahaan yang sama, dan [9] menemukan bahwa pentingnya rasio keuangan bersifat kontekstual terhadap metode analisis. Dalam literatur seleksi fitur, stabilitas didefinisikan sebagai konsistensi peringkat atau subset fitur ketika data latih berubah, dan umum diukur dengan korelasi peringkat serta indeks Jaccard [21].

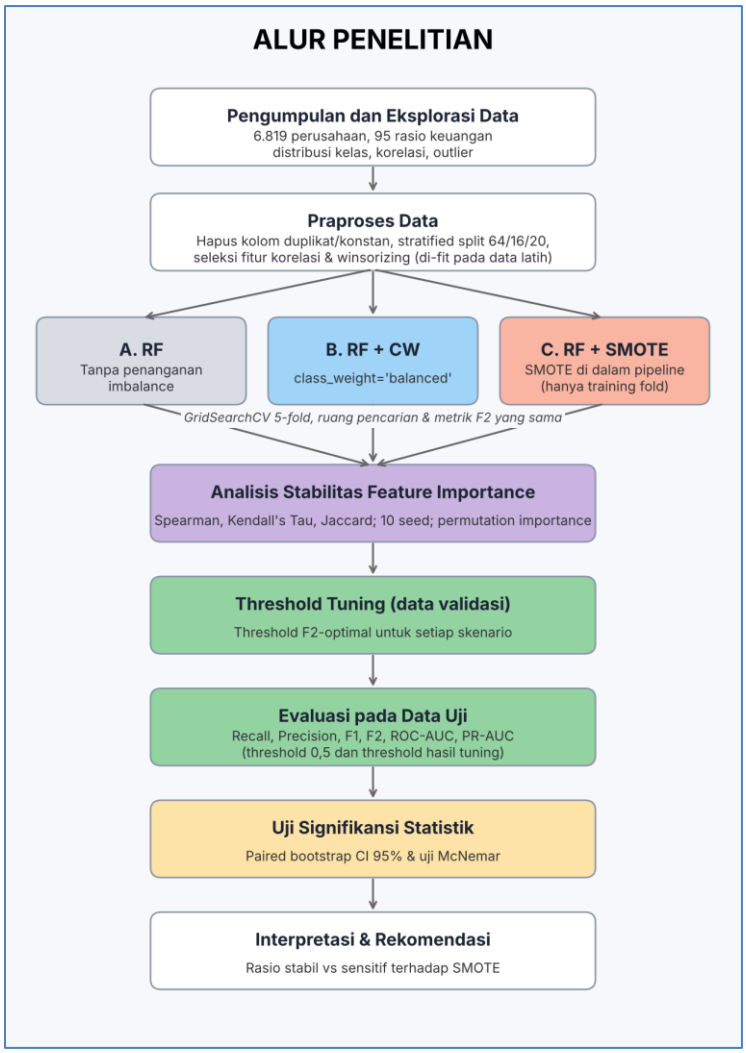
Berdasarkan uraian tersebut, penelitian ini menguji stabilitas peringkat *feature importance Random Forest* sebelum dan sesudah SMOTE menggunakan korelasi Spearman, Kendall's Tau, dan kesamaan Jaccard, dipadukan dengan evaluasi berorientasi *Recall* sebagaimana ditekankan [3] dan [4]. Untuk menjawab catatan metodologis di atas, SMOTE diterapkan di dalam *pipeline cross-validation*, efek SMOTE dipisahkan dari efek pembobotan kelas dan *threshold tuning*, serta stabilitas diuji lintas beberapa *random seed*.

 ISSN:2302-8149
e-ISSN:2540-9719

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan desain komparatif untuk menganalisis dampak penerapan SMOTE terhadap model *Random Forest* dalam memprediksi *financial distress* perusahaan berdasarkan rasio keuangan. Dalam penelitian ini, *financial distress* dioperasionalkan sebagai status kebangkrutan perusahaan (variabel target *Bankrupt?*) sebagaimana didefinisikan pada sumber data [15]; dengan demikian, hasil penelitian lebih tepat dibaca sebagai prediksi kebangkrutan, yaitu bentuk *financial distress* yang paling berat. *Random Forest* dipilih karena mampu menangani data berdimensi tinggi, relatif tahan terhadap *overfitting* melalui mekanisme *bagging*, dan menyediakan ukuran *feature importance* untuk mengidentifikasi rasio keuangan yang paling berkontribusi terhadap prediksi [18].

Alur penelitian diawali dengan pengumpulan data sekunder dari platform Kaggle dan *Exploratory Data Analysis* (EDA). Praproses data mencakup penghapusan kolom duplikat dan konstan, pembagian data latih, validasi, dan uji, seleksi fitur berbasis korelasi, serta penanganan *outlier*. Tiga skenario *Random Forest* kemudian dibangun dengan ruang pencarian *hyperparameter* yang sama, yaitu tanpa penanganan ketidakseimbangan (RF), dengan pembobotan kelas (RF + CW), dan dengan SMOTE (RF + SMOTE). Selanjutnya dilakukan analisis stabilitas *feature importance*, optimasi ambang batas klasifikasi, evaluasi pada data uji, dan uji signifikansi statistik. Tahapan penelitian disajikan pada Gambar 1 dan diuraikan pada subbagian 3.1 sampai 3.8.



Gambar 1 Alur penelitian

 ISSN:2302-8149
e-ISSN:2540-9719

3.1 Pengumpulan Data

Data yang digunakan adalah data sekunder *Company Bankruptcy Prediction Dataset* yang diunduh dari platform Kaggle (diunggah oleh fedesoriano, 2021). Data bersumber dari *Taiwan Economic Journal* (TEJ) untuk perusahaan di Taiwan pada periode 1999–2009 dan merupakan data yang digunakan pada penelitian [15], dengan status kebangkrutan ditentukan berdasarkan regulasi bisnis Bursa Efek Taiwan. Dataset terdiri atas 6.819 baris (perusahaan) dan 96 kolom, yaitu 95 rasio keuangan sebagai variabel prediktor dan satu kolom target biner (*Bankrupt?*). Sebanyak 220 perusahaan (3,23%) berstatus bangkrut, sehingga data bersifat *highly imbalanced* dan relevan untuk menguji dampak penanganan ketidakseimbangan kelas.

Eksplorasi data (*Exploratory Data Analysis*/EDA) dilakukan setelah data sekunder terkumpul untuk memahami karakteristik data sebelum pemodelan. Eksplorasi mencakup pemeriksaan struktur dan tipe data, pengecekan nilai hilang (*missing value*) dan data duplikat, analisis distribusi variabel target untuk mengidentifikasi tingkat ketidakseimbangan kelas, analisis statistik deskriptif variabel prediktor, analisis korelasi antar fitur dan korelasi fitur terhadap target menggunakan koefisien Spearman, serta deteksi *outlier* menggunakan metode *Interquartile Range* (IQR). Hasil eksplorasi digunakan sebagai dasar dalam menentukan keputusan pra proses data pada tahap berikutnya. Tahap EDA bersifat informatif dan tidak melibatkan penghapusan maupun perubahan terhadap data.

3.2 Praproses Data (*Data Preprocessing*)

Praproses data dilakukan melalui beberapa tahapan. Kolom yang identik secara struktural (duplikat) dan kolom tanpa variansi (konstan) dihapus untuk menghilangkan informasi redundan. Data kemudian dibagi menjadi data latih (64%), data validasi (16%), dan data uji (20%) menggunakan *stratified split* dengan *random_state*=42 sehingga proporsi kelas bangkrut terjaga pada setiap subset. Seleksi fitur dilakukan dengan menghapus salah satu dari setiap pasangan fitur yang memiliki korelasi Spearman absolut lebih dari 0,8, yaitu fitur yang korelasinya terhadap target lebih rendah. Nilai ekstrem pada empat fitur dengan *outlier* terbanyak ditangani dengan *winsorizing* pada persentil ke-1 dan ke-99. Seluruh langkah yang memerlukan estimasi dari data (korelasi untuk seleksi fitur dan batas *winsorizing*) dihitung hanya dari data latih, lalu diterapkan dengan aturan yang sama pada data validasi dan data uji untuk mencegah kebocoran informasi (*data leakage*).

3.3 Pelatihan Model: Tiga Skenario Pembanding

Tiga skenario *Random Forest* dilatih menggunakan ruang pencarian *hyperparameter* yang sama melalui GridSearchCV dengan validasi silang 5-fold berstratifikasi (StratifiedKFold, *shuffle*=True, *random_state*=42). Ruang pencarian yang sama memastikan setiap skenario memperoleh kesempatan optimasi yang setara, tetapi konfigurasi *hyperparameter* terbaik dapat berbeda antar-skenario karena ditentukan oleh data latih masing-masing. Seluruh model menggunakan *random_state*=42 dan *n_jobs*=-1. Ketiga skenario adalah sebagai berikut.

- **Skenario A – RF:** *Random Forest* dilatih pada data latih asli tanpa penanganan ketidakseimbangan kelas (*class_weight*=None). Skenario ini merupakan kontrol untuk mengisolasi dampak SMOTE.

- **Skenario B – RF + CW:** *Random Forest* dengan *class_weight*='balanced', yaitu penanganan ketidakseimbangan melalui pembobotan kelas tanpa mengubah distribusi data. Skenario ini digunakan untuk menilai apakah SMOTE memberikan manfaat melebihi pembobotan kelas.

- **Skenario C – RF + SMOTE:** SMOTE (*sampling_strategy*='auto' sehingga jumlah kelas minoritas disamakan 1:1 dengan kelas mayoritas, *k_neighbors*=5, *random_state*=42) dikombinasikan dengan *Random Forest* tanpa *class_weight*. Pembobotan kelas tidak digunakan karena setelah SMOTE distribusi kelas sudah seimbang, sehingga pembobotan tidak lagi berpengaruh dan hanya akan mengaburkan efek SMOTE.

Untuk mencegah *data leakage*, SMOTE ditempatkan sebagai langkah di dalam *pipeline* (imblearn.pipeline.Pipeline) yang dievaluasi oleh GridSearchCV. Dengan susunan ini, pada setiap

 ISSN:2302-8149
e-ISSN:2540-9719

iterasi *cross-validation* SMOTE hanya diterapkan pada *training fold*, sedangkan *validation fold* tetap berisi data asli; data validasi dan data uji juga tidak pernah di-*resample* [20]. Setiap skenario dioptimasi menggunakan F2-Score sebagai metrik skoring karena memberi bobot lebih besar pada *Recall* dibandingkan *Precision*. Ruang pencarian meliputi n_estimators {100, 200, 300}, max_depth {None, 10, 20}, min_samples_leaf {1, 2, 4}, dan max_features {sqrt, log2, 0,5}, sehingga terdapat 81 kombinasi untuk setiap skenario. Eksperimen dijalankan menggunakan Python 3.13.16, scikit-learn 1.9.1, imbalanced-learn 0.14.2, pandas 3.0.5, NumPy 2.5.3, dan SciPy 1.18.1.

3.4 Analisis Stabilitas Feature Importance

Nilai *feature importance* berbasis MDI diekstraksi dari model terbaik setiap skenario. Perbandingan utama dilakukan antara skenario RF dan RF + SMOTE untuk mengisolasi dampak SMOTE, sedangkan perbandingan RF + CW dan RF + SMOTE dilaporkan sebagai pelengkap. Kesesuaian peringkat diukur dengan korelasi Spearman dan Kendall's Tau pada seluruh fitur, serta kesamaan Jaccard pada kelompok Top-5, Top-10, dan Top-15 (rumus (6)–(8) pada subbagian 3.6). Karena *Random Forest* bersifat stokastik, setiap skenario juga dilatih ulang dengan 10 *random seed* (0–9) menggunakan *hyperparameter* terbaiknya. Kemiripan peringkat antar-seed dalam satu skenario (*within-scenario*) dijadikan acuan variasi alami, lalu dibandingkan dengan kemiripan antara RF dan RF + SMOTE pada seed yang sama (*between-scenario*) menggunakan uji Mann-Whitney satu sisi. Pergeseran peringkat dinyatakan disebabkan oleh SMOTE hanya jika lebih besar daripada variasi alami tersebut. Sebagai pembanding MDI, dihitung pula *permutation importance* pada data validasi dengan metrik *Average Precision* dan 10 pengulangan [23].

3.5 Optimasi Ambang Batas Klasifikasi (*Threshold Tuning*)

Ambang batas klasifikasi dioptimasi pada data validasi dengan memaksimalkan F2-Score untuk setiap skenario. Model kemudian dievaluasi pada data uji menggunakan *threshold* bawaan 0,5 dan *threshold* hasil *tuning*, sehingga efek SMOTE (perbandingan antar-skenario pada *threshold* yang sama) dapat dipisahkan dari efek *threshold tuning* (perbandingan dalam skenario yang sama). Rumus F2-Score disajikan pada subbagian 3.6.

3.6 Metrik Evaluasi Model

Kinerja model diukur melalui *accuracy*, *precision*, *recall*, F1-Score, F2-Score, ROC-AUC, dan PR-AUC (*Average Precision*). Stabilitas peringkat *feature importance* diukur dengan korelasi Spearman, Kendall's Tau, dan kesamaan Jaccard. *Confusion matrix* digunakan untuk memvisualisasikan sebaran prediksi benar dan salah pada setiap kelas.

Accuracy merupakan rasio antara jumlah prediksi yang benar terhadap keseluruhan data uji. Metrik ini memberikan gambaran umum performa model, namun dapat menyesatkan ketika data memiliki distribusi kelas yang tidak seimbang. Secara matematis, akurasi dinyatakan dalam rumus (1):

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

di mana *TP (True Positive)* adalah jumlah prediksi benar untuk kelas positif, *TN (True Negative)* adalah jumlah prediksi benar untuk kelas negatif, *FP (False Positive)* adalah jumlah prediksi salah untuk kelas positif, dan *FN (False Negative)* adalah jumlah prediksi salah untuk kelas negatif.

Precision mengukur seberapa tepat model dalam memprediksi kelas positif. Rumus (2) mendefinisikan *precision* sebagai berikut:

$$\text{Precision}=\frac{TP}{TP+FP} \quad (2)$$

Recall mengukur seberapa lengkap model dalam menangkap seluruh kelas positif yang sebenarnya. *Recall* dinyatakan dalam rumus (3):

$$\text{Recall}=\frac{TP}{TP+FN} \quad (3)$$

 ISSN:2302-8149
e-ISSN:2540-9719

F1-Score merupakan rata-rata harmonik dari *precision* dan *recall*. Rumus (4) mendefinisikan F1-Score sebagai berikut:

$$\text{F1-Score} = 2\ X\ \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

F2-Score merupakan generalisasi dari F1-Score yang memberikan bobot dua kali lebih besar pada *recall* dibandingkan *precision*. Metrik ini dipilih sebagai kriteria utama optimasi model (pada GridSearchCV dan *threshold tuning*) karena sejalan dengan tujuan penelitian untuk meminimalkan *False Negative* pada deteksi *financial distress*. Rumus (5) mendefinisikan F2-Score sebagai berikut:

$$\text{Fβ-Score} = (1 + \beta^2) \times \frac{Precision \times Recall}{(\beta^2 \times Precision) + Recall}, \beta = 2 \quad (5)$$

Korelasi Spearman (ρ) digunakan untuk mengukur kesesuaian peringkat *feature importance* antara dua model. Rumus (6) mendefinisikan korelasi Spearman sebagai berikut:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \quad (6)$$

di mana d_i adalah selisih peringkat fitur ke-i antara kedua model dan n adalah jumlah fitur yang dibandingkan. Nilai ρ mendekati 1 menunjukkan urutan peringkat sangat mirip, sedangkan nilai mendekati 0 menunjukkan tidak ada kesesuaian urutan. Nilai *p* pada uji korelasi hanya menunjukkan bahwa korelasi berbeda dari nol, sehingga tidak digunakan sebagai bukti stabilitas; stabilitas dinilai dari besaran koefisien yang dibandingkan dengan variasi antar-seed.

Korelasi Kendall's Tau (τ) digunakan sebagai ukuran pelengkap untuk menilai konkordansi peringkat antar kedua model. Rumus (7) mendefinisikan Kendall's Tau sebagai berikut:

$$T = \frac{(C-D)}{\frac{1}{2}n(n-1)} \quad (7)$$

di mana *C* adalah jumlah pasangan konkordan, *D* adalah jumlah pasangan diskordan, dan *n* adalah jumlah fitur.

Kesamaan Jaccard digunakan untuk mengukur proporsi kesamaan anggota fitur pada kelompok Top-K antara kedua model, tanpa memperhatikan urutan peringkat di dalamnya. Rumus (8) mendefinisikan kesamaan Jaccard sebagai berikut:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (8)$$

di mana *A* adalah himpunan Top-*K* fitur terpenting pada model pertama dan *B* adalah himpunan Top-*K* fitur terpenting pada model kedua. Nilai 1 berarti kedua himpunan identik. Karena penyebutnya adalah gabungan kedua himpunan, nilai Jaccard tidak sama dengan persentase fitur yang sama; sebagai contoh, irisan 9 dari 10 fitur menghasilkan J = 9/11 = 0,818.

3.7 Pengujian Model dan Uji Signifikansi Statistik

Model terbaik setiap skenario diuji pada data uji yang tidak dilibatkan dalam pelatihan, pemilihan *hyperparameter*, maupun *threshold tuning*. Evaluasi dilakukan pada enam kondisi, yaitu tiga skenario masing-masing dengan *threshold* 0,5 dan *threshold* hasil *tuning* F2 (subbagian 3.5).

Signifikansi perbedaan performa dinilai dengan dua uji. Pertama, *paired bootstrap* dengan 2.000 replikasi pada data uji (pengambilan sampel ulang berstratifikasi per kelas) untuk memperoleh selang kepercayaan (CI) 95% dari selisih *Recall*, *Precision*, F2-Score, ROC-AUC, dan PR-AUC; perbedaan dinyatakan signifikan apabila CI tidak memuat nol. Kedua, uji McNemar *exact* [22] untuk membandingkan pola kesalahan klasifikasi dua model pada data uji yang sama, baik pada seluruh data

 ISSN:2302-8149
e-ISSN:2540-9719

maupun khusus pada perusahaan bangkrut. Istilah "signifikan" dalam naskah ini hanya digunakan untuk perbedaan yang lolos uji tersebut pada $\alpha = 0{,}05$.

3.8 Interpretasi dan Kontribusi Praktis

Sebagai tahap akhir, rasio keuangan dikelompokkan berdasarkan hasil pengulangan 10 *seed* (subbagian 3.4) menjadi dua kategori: (1) **stabil dan penting**, yaitu rasio yang masuk Top-10 pada minimal 80% *seed* di kedua skenario RF dan RF + SMOTE; dan (2) **sensitif terhadap SMOTE**, yaitu rasio dengan selisih peringkat rata-rata minimal 10 posisi dan lebih besar dari dua kali simpangan baku peringkat antar-seed. Kriteria kedua memastikan bahwa pergeseran yang dilaporkan bukan sekadar akibat keacakan model.

4 Hasil dan Pembahasan

4.1 Hasil Pengumpulan Data

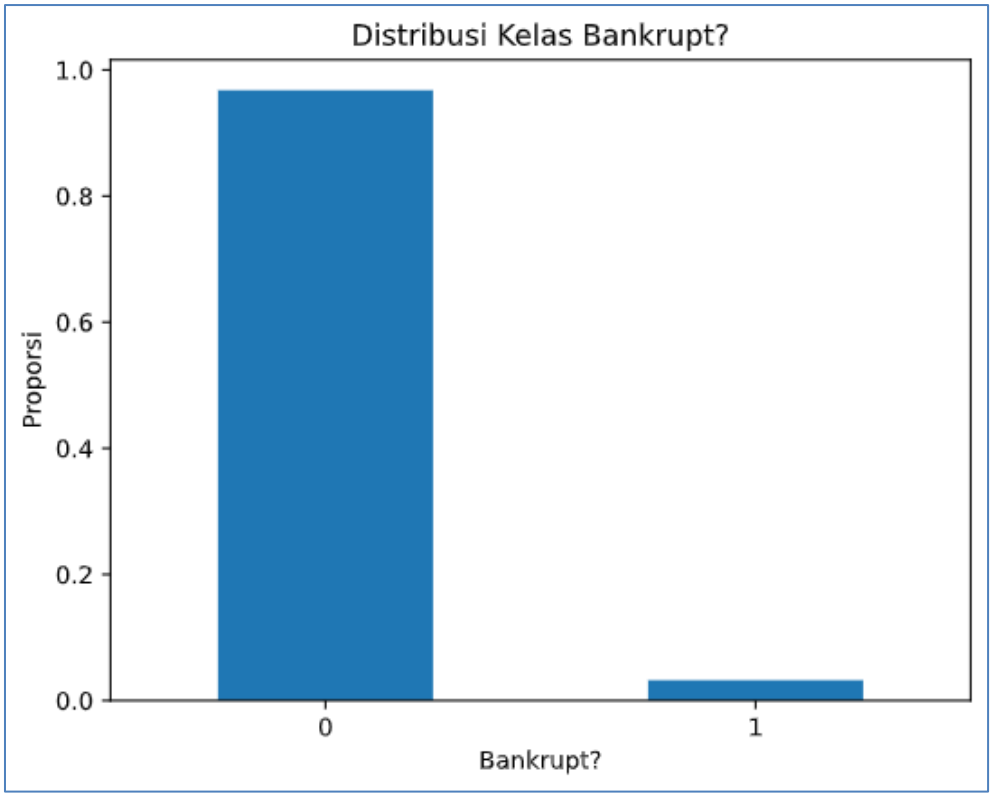
Dataset yang dimuat terdiri atas 6.819 baris dan 96 kolom, yaitu 95 rasio keuangan dan satu kolom target (*Bankrupt?*), sesuai spesifikasi dataset. Seluruh kolom bertipe numerik (float64/int64) tanpa *missing value* dan tanpa baris duplikat, sehingga tidak diperlukan imputasi maupun *encoding* kategori (Gambar 2).

	Bankrupt?	ROA(C) before interest and depreciation before interest	ROA(A) before interest and % after tax	ROA(B) before interest and depreciation after tax	Operating Gross Margin	Realized Sales Gross Margin	Operating Profit Rate	Pre-tax net Interest Rate	After-tax net Interest Rate	Non-industry income and expenditure/revenue	Continuous interest rate (after tax)	Operating Expense Rate	Research and development expense rate	Cash flow rate
0	1	0.370594	0.424389	0.405750	0.601457	0.601457	0.998969	0.796887	0.808809	0.302646	0.780985	1.256969e-04	0.0	0.458143
1	1	0.464291	0.538214	0.516730	0.610235	0.610235	0.998946	0.797380	0.809301	0.303556	0.781506	2.897851e-04	0.0	0.461867
2	1	0.426071	0.499019	0.472295	0.601450	0.601364	0.998857	0.796403	0.808388	0.302035	0.780284	2.361297e-04	25500000.0	0.458521
3	1	0.399844	0.451265	0.457733	0.583541	0.583541	0.998700	0.796967	0.808966	0.303350	0.781241	1.078888e-04	0.0	0.465705
4	1	0.465022	0.538432	0.522298	0.598783	0.598783	0.998973	0.797366	0.809304	0.303475	0.781550	7.890000e+09	0.0	0.462746

Gambar 2 Cuplikan dataset

Distribusi variabel target menunjukkan bahwa 6.599 perusahaan (96,77%) berstatus tidak bangkrut dan 220 perusahaan (3,23%) berstatus bangkrut (Gambar 3). Kondisi *highly imbalanced* ini menjelaskan mengapa akurasi tidak memadai sebagai metrik utama: model yang memprediksi seluruh perusahaan sebagai "tidak bangkrut" tetap memperoleh akurasi di atas 96% tanpa mendeteksi satu pun kasus kebangkrutan.

 ISSN:2302-8149
e-ISSN:2540-9719

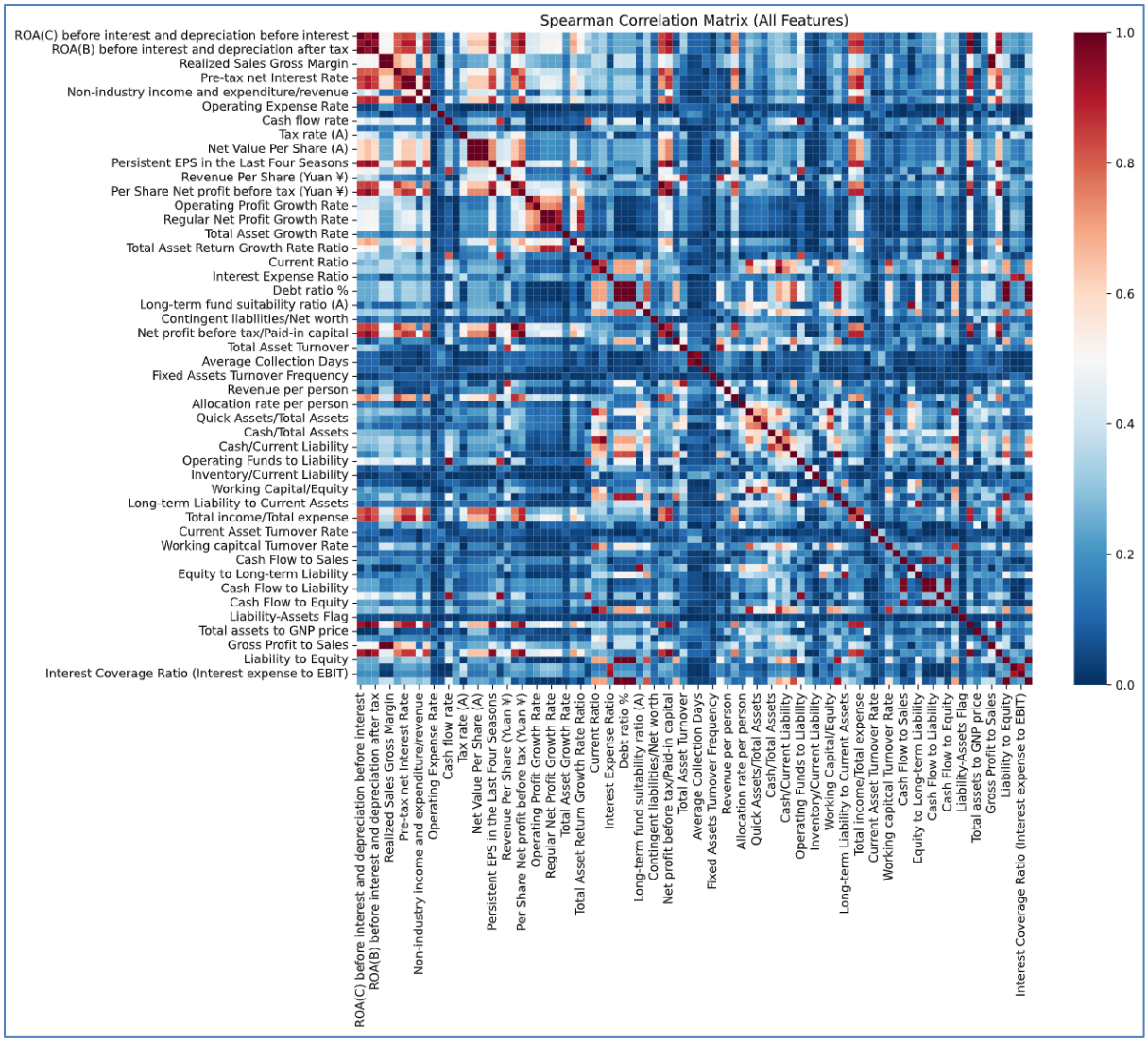


Gambar 3 Distribusi kelas target

Statistik deskriptif variabel prediktor mengungkap rentang nilai yang sangat bervariasi antar rasio keuangan yakni sebagian fitur berskala rasio kecil (0–1), sementara sebagian lain (seperti *Revenue Per Share*) berskala jutaan hingga milyaran. Oleh karena itu, standarisasi/normalisasi skala tidak dilakukan pada penelitian ini, mengingat algoritma Random Forest bersifat *scale-invariant* yaitu pembagian node pada setiap pohon keputusan didasarkan pada ambang batas nilai fitur, bukan jarak antar titik data, sehingga tidak terpengaruh oleh perbedaan skala antar fitur.

Analisis korelasi Spearman antarfitur (Gambar 4) menunjukkan sejumlah kelompok fitur dengan korelasi sangat tinggi (mendekati 1,0), terutama pada rasio profitabilitas seperti *ROA(C) before interest and depreciation before interest*, *ROA(B) before interest and depreciation after tax*, dan *Net profit before tax/Paid-in capital*, yang secara konseptual berasal dari komponen laba yang serupa. Temuan multikolinearitas ini menjadi dasar seleksi fitur berbasis korelasi pada tahap praproses (subbagian 3.2 dan 4.2).

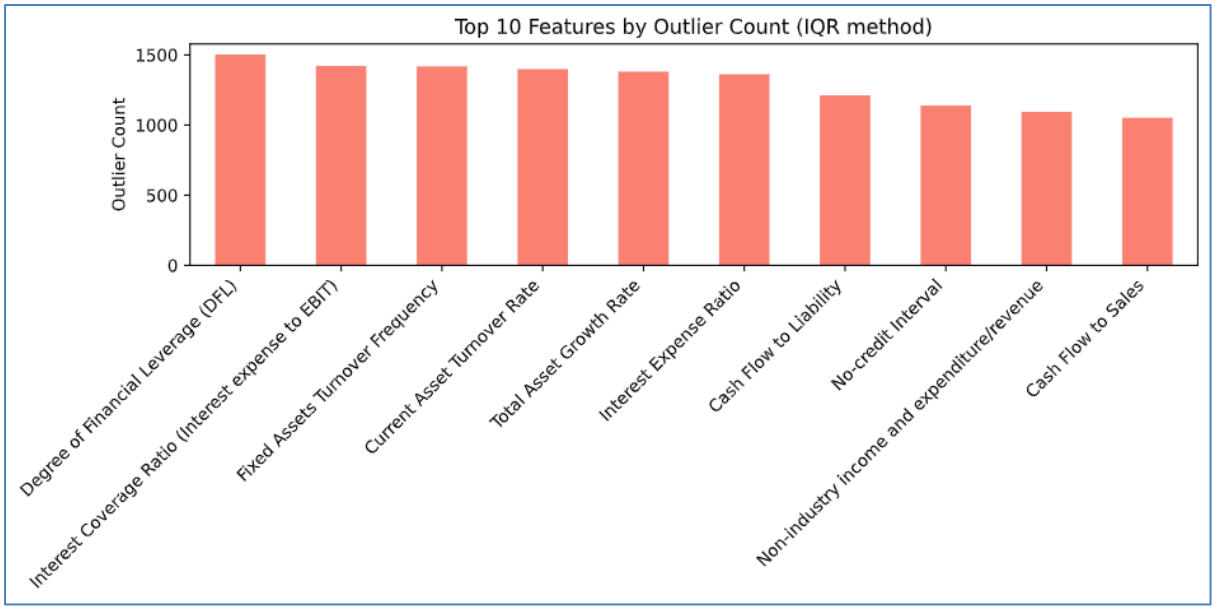
ISSN:2302-8149
e-ISSN:2540-9719



Gambar 4 Heatmap korelasi spearman antarfitur

Deteksi *outlier* dengan metode *Interquartile Range* (IQR) (Gambar 5) menunjukkan bahwa 87 dari 92 fitur memiliki *outlier*. Jumlah terbanyak terdapat pada *Degree of Financial Leverage* (DFL) dengan 1.503 observasi, diikuti *Interest Coverage Ratio*, *Fixed Assets Turnover Frequency*, *Current Asset Turnover Rate*, dan *Total Asset Growth Rate*. Banyaknya *outlier* pada rasio leverage dan perputaran wajar terjadi pada data laporan keuangan karena rasio tersebut sensitif terhadap kondisi ekstrem perusahaan, misalnya laba operasi yang mendekati nol.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 5 Top-10 fitur dengan jumlah outlier terbanyak

Secara keseluruhan, tahap EDA ini mengarahkan tiga keputusan penting pada tahap pra proses data: (1) tidak diperlukannya standarisasi skala karena karakteristik algoritma Random Forest; (2) perlunya seleksi fitur berbasis korelasi untuk mengatasi multikolinearitas yang teridentifikasi pada heatmap; dan (3) perlunya penanganan *outlier* melalui *winsorizing* pada fitur-fitur dengan proporsi nilai ekstrim tertinggi.

4.2 Hasil Praproses Data

Penghapusan kolom duplikat struktural (*Current Liability to Liability* dan *Current Liability to Equity*) serta kolom konstan (*Net Income Flag*) menghasilkan 92 fitur. Pembagian data secara *stratified* menghasilkan 4.364 baris data latih (141 bangkrut), 1.091 baris data validasi (35 bangkrut), dan 1.364 baris data uji (44 bangkrut), dengan proporsi kelas bangkrut sekitar 3,2% pada ketiga subset. Jumlah kasus bangkrut yang kecil pada data validasi dan data uji perlu diperhatikan karena membuat estimasi metrik untuk kelas minoritas memiliki ketidakpastian yang cukup besar.

Seleksi fitur berbasis korelasi dihitung pada data latih: dari setiap pasangan fitur dengan korelasi Spearman absolut lebih dari 0,8, fitur dengan korelasi lebih rendah terhadap target dihapus. Sebanyak 53 fitur dihapus (Gambar 6), sehingga tersisa 39 fitur untuk pemodelan. Perlu dicatat bahwa beberapa rasio yang secara konseptual penting, seperti ROA dan *Current Ratio*, ikut terhapus karena diwakili oleh fitur lain yang sangat berkorelasi dengannya; interpretasi *feature importance* pada bagian selanjutnya harus dibaca sebagai kepentingan fitur representatif dari kelompok rasio tersebut.

 ISSN:2302-8149
e-ISSN:2540-9719

```python
def remove_highly_correlated_features(X_train, y_train, threshold=0.8):
    df_train = X_train.copy()
    df_train['Bankrupt?'] = y_train

    corr_matrix = X_train.corr(method='spearman').abs()
    corr_with_target = df_train.corr(method='spearman')['Bankrupt?'].abs().drop('Bankrupt?')

    upper = corr_matrix.where(np.triu(np.ones(corr_matrix.shape), k=1).astype(bool))
    to_drop = set()

    for col in upper.columns:
        high_corr = upper.index[upper[col] > threshold].tolist()
        for feature in high_corr:
            if corr_with_target[feature] < corr_with_target[col]:
                to_drop.add(feature)
            else:
                to_drop.add(col)

    return sorted(to_drop)

to_drop = remove_highly_correlated_features(X_train, y_train, threshold=0.8)
print(f'Features to drop ({len(to_drop)}):')
for c in to_drop:
    print(f'  - {c}')
```

```
Features to drop (53):
  - Accounts Receivable Turnover
  - After-tax Net Profit Growth Rate
  - After-tax net Interest Rate
  - Borrowing dependency
  - CFO to Assets
  - Cash Flow Per Share
  - Cash Flow to Liability
  - Cash Flow to Sales
  - Cash Flow to Total Assets
  - Cash Reinvestment %
  - Cash flow rate
  - Cash/Current Liability
  - Continuous interest rate (after tax)
  - Current Assets/Total Assets
  - Current Liabilities/Equity
  - Current Liability to Assets
  - Current Liability to Current Assets
  - Current Ratio
  - Debt ratio %
  - Equity to Liability
  - Fixed Assets to Assets
  - Gross Profit to Sales
```

Gambar 6. Daftar Fitur yang Dihapus Akibat Korelasi Tinggi

Winsorizing pada persentil ke-1 dan ke-99 diterapkan pada empat fitur dengan *outlier* paling ekstrem, yaitu DFL, *Fixed Assets Turnover Frequency*, *Current Asset Turnover Rate*, dan *Total Asset Growth Rate*. Batas persentil dihitung hanya dari data latih lalu diterapkan pada data validasi dan data uji. Karena *Random Forest* memisahkan *node* berdasarkan urutan nilai, langkah ini terutama membatasi nilai yang sangat ekstrem dan tidak mengubah urutan sebagian besar observasi.

Hasil akhir praproses adalah 39 fitur yang digunakan secara identik pada ketiga skenario model.

4.3 Hasil Pelatihan Model: Tiga Skenario

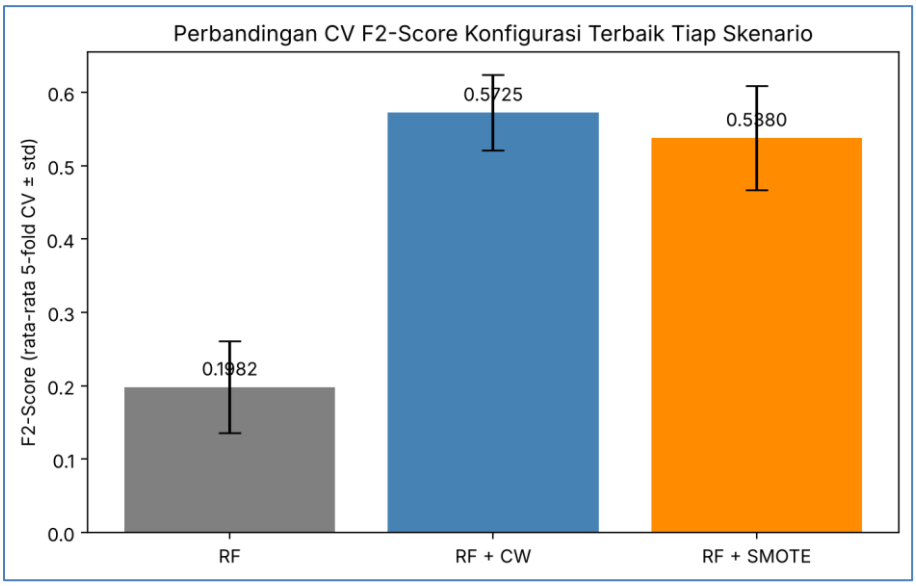
Ketiga skenario dioptimasi dengan ruang pencarian *hyperparameter*, skema 5-fold *cross-validation*, dan metrik F2-Score yang sama. Hasil konfigurasi terbaik setiap skenario disajikan pada Tabel 1 dan Gambar 7. Verifikasi *pipeline* menunjukkan bahwa pada setiap fold, SMOTE hanya memperbesar *training fold* (misalnya 3.491 menjadi 6.756 baris, dengan kelas minoritas dari 113 menjadi 3.378), sedangkan *validation fold* tetap berisi 872–873 baris data asli.

Tanpa penanganan ketidakseimbangan, skenario RF hanya mencapai CV F2-Score 0,1982, karena pada *threshold* 0,5 model sangat jarang memprediksi kelas bangkrut. Penanganan ketidakseimbangan meningkatkan CV F2-Score secara nyata, baik melalui pembobotan kelas (RF + CW: 0,5725) maupun SMOTE (RF + SMOTE: 0,5380). Selisih antara RF + CW dan RF + SMOTE lebih kecil daripada simpangan baku antar-fold, sehingga kedua pendekatan memberikan hasil optimasi yang sebanding.

Tabel 1 Hasil GridSearchCv (rata-rata f2-score 5-fold ± simpangan baku) dan hyperparameter terbaik

Skenario	CV F2-Score	n_estimators	max_depth	min_samples_leaf	max_features
RF	0,1982 ± 0,0624	300	None	1	0,5
RF + CW	0,5725 ± 0,0515	100	None	4	log2
RF + SMOTE	0,5380 ± 0,0710	200	10	4	log2

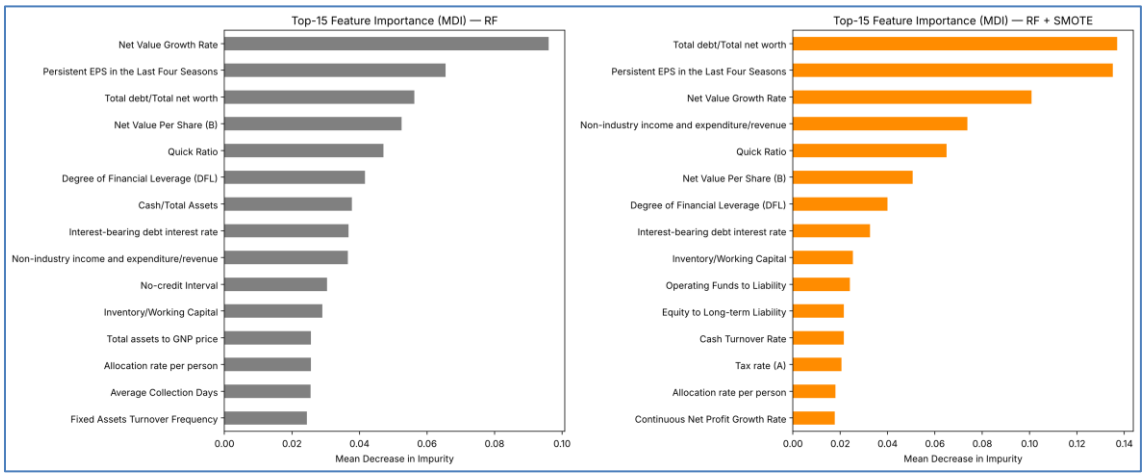
Konfigurasi terbaik ketiga skenario berbeda (Tabel 1), padahal ruang pencariannya sama. Skenario dengan penanganan ketidakseimbangan cenderung memilih min_samples_leaf=4 dan max_features='log2', yaitu pohon yang lebih teregularisasi, sedangkan RF memilih daun sekecil mungkin agar dapat mengisolasi sedikit kasus bangkrut. Karena itu, perbedaan hasil antar-skenario mencerminkan dampak keseluruhan perlakuan data beserta *hyperparameter* optimal yang menyertainya, bukan perlakuan data secara terisolasi; hal ini menjadi keterbatasan desain yang dibahas pada subbagian 4.8.



Gambar 7 Perbandingan CV F2-score konfigurasi terbaik tiap skenario

4.4 Analisis Stabilitas Feature Importance

Nilai *feature importance* MDI diekstraksi dari model terbaik ketiga skenario. Gambar 8 memperlihatkan 15 fitur teratas pada model RF dan RF + SMOTE, sedangkan Tabel 2 merangkum ukuran kesesuaian peringkat antar-skenario.



Gambar 8. Top-15 Feature Importance (MDI): RF vs RF + SMOTE

 ISSN:2302-8149
e-ISSN:2540-9719

Pada perbandingan utama RF vs RF + SMOTE, korelasi Spearman sebesar 0,6148 dan Kendall's Tau 0,4602 menunjukkan kesesuaian peringkat yang moderat. Kesamaan Jaccard Top-10 sebesar 0,667 berarti 8 dari 10 fitur teratas sama pada kedua model (8/12 = 0,667; nilai Jaccard bukan persentase fitur yang sama). Pada level teratas, *Total debt/Total net worth* dan *Persistent EPS in the Last Four Seasons* tetap berada pada peringkat 1–3 di kedua model, sementara beberapa fitur di peringkat menengah bergeser cukup jauh (Gambar 8).

Tabel 2 Kesesuaian peringkat feature importance antar-skenario (model terbaik, seed 42)

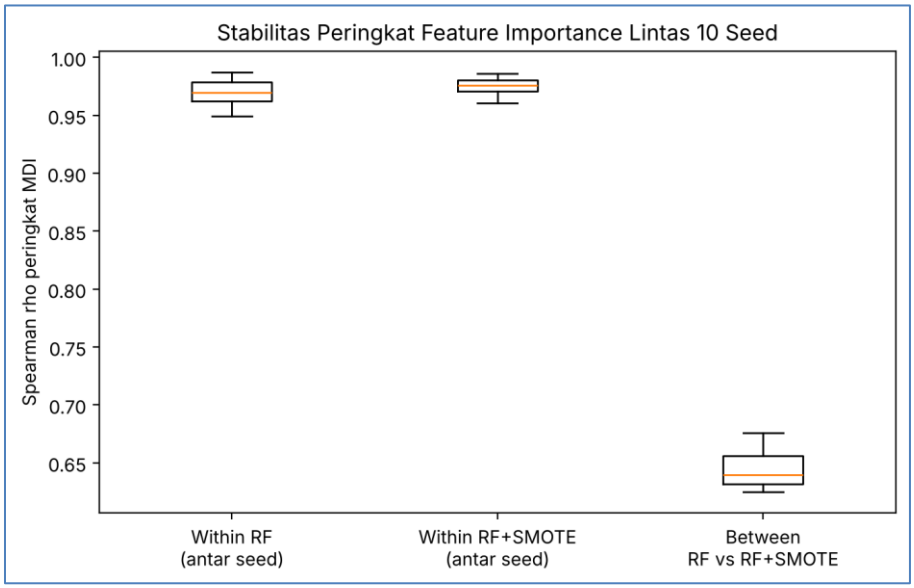
perbandingan	spearman ρ	kendall's τ	jaccard top-5 (irisan)	jaccard top-10 (irisan)	jaccard top-15 (irisan)
RF vs RF + SMOTE	0,615	0,460	0,667 (4/5)	0,667 (8/10)	0,500 (10/15)
RF + CW vs RF + SMOTE	0,867	0,722	1,000 (5/5)	0,818 (9/10)	0,765 (13/15)
RF vs RF + CW	0,646	0,484	0,667 (4/5)	0,667 (8/10)	0,579 (11/15)

Perbandingan RF + CW vs RF + SMOTE menghasilkan kesesuaian yang jauh lebih tinggi (Spearman 0,8668; Jaccard Top-10 0,818 atau irisan 9/10). Sebaliknya, RF vs RF + CW memiliki kesesuaian yang serupa dengan RF vs RF + SMOTE (Spearman 0,6457). Pola ini menunjukkan bahwa sebagian besar pergeseran peringkat disebabkan oleh penyeimbangan kelas itu sendiri, baik melalui pembobotan maupun sampel sintetis, bukan semata-mata oleh proses interpolasi SMOTE. Nilai *p* dari uji korelasi tidak digunakan sebagai bukti stabilitas karena hanya menguji apakah korelasi berbeda dari nol.

Untuk memastikan bahwa pergeseran tersebut bukan akibat keacakan *Random Forest*, setiap skenario dilatih ulang dengan 10 *random seed* (Tabel 3 dan Gambar 9). Dalam satu skenario, peringkat MDI sangat konsisten antar-seed (Spearman 0,970 untuk RF dan 0,975 untuk RF + SMOTE), sedangkan kesesuaian antara RF dan RF + SMOTE hanya 0,644 ± 0,017. Uji Mann-Whitney menunjukkan bahwa kesesuaian *between-scenario* lebih rendah secara signifikan daripada *within-scenario* (U = 0; p < 0,001). Dengan demikian, SMOTE memang mengubah peringkat *feature importance* secara moderat dan perubahan tersebut melampaui variasi acak model.

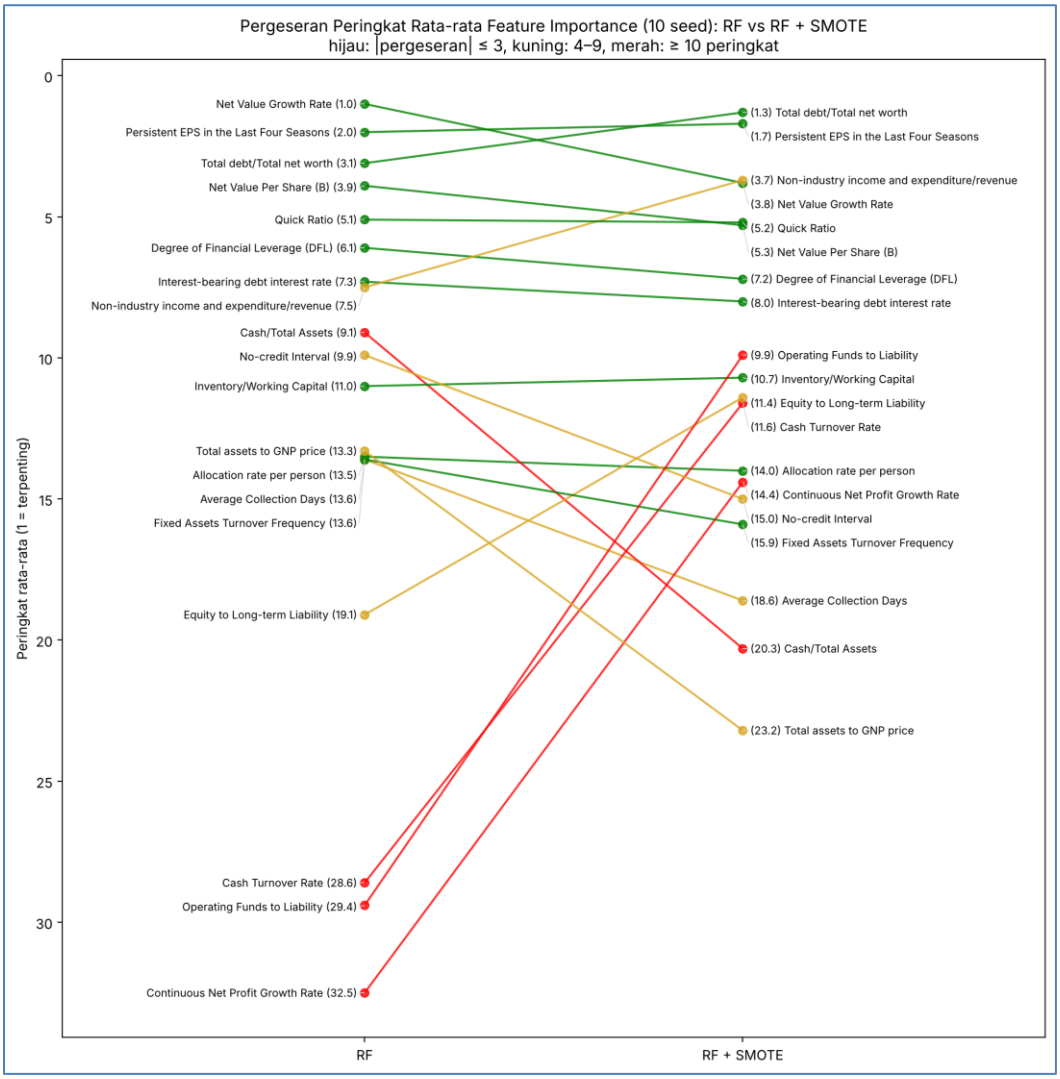
Tabel 3 Stabilitas peringkat feature importance lintas 10 random seed (rata-rata ± simpangan baku)

Perbandingan	Spearman ρ	Kendall's τ	Jaccard Top-10
Within RF (antar-seed)	0,970 ± 0,010	0,890 ± 0,021	1,000 ± 0,000
Within RF + CW (antar-seed)	0,948 ± 0,015	0,829 ± 0,026	0,785 ± 0,136
Within RF + SMOTE (antar-seed)	0,975 ± 0,007	0,885 ± 0,019	0,805 ± 0,100
Between RF vs RF + SMOTE	0,644 ± 0,017	0,481 ± 0,015	0,682 ± 0,048
Between RF + CW vs RF + SMOTE	0,852 ± 0,031	0,694 ± 0,037	0,699 ± 0,091
Between RF vs RF + CW	0,686 ± 0,028	0,525 ± 0,037	0,684 ± 0,081



Gambar 9 Distribusi korelasi spearman peringkat MDI lintas 10 seed

Gambar 10 memperlihatkan pergeseran peringkat rata-rata (10 *seed*) setiap fitur dari RF ke RF + SMOTE. Garis hijau menandai rasio yang stabil (bergeser maksimal 3 peringkat), garis kuning bergeser 4–9 peringkat, dan garis merah bergeser minimal 10 peringkat.



Gambar 10 Pergeseran peringkat rata-rata feature importance (10 seed): RF vs RF + SMOTE

Sebagai pembanding MDI, *permutation importance* pada data validasi menghasilkan kesesuaian yang jauh lebih rendah antara RF dan RF + SMOTE (Spearman -0,082; irisan Top-10 2/10). Kesesuaian MDI dengan *permutation importance* juga hanya moderat pada RF (Spearman 0,517; irisan Top-10 7/10) dan rendah pada RF + SMOTE (Spearman -0,227; irisan 2/10). Sebagian ketidaksesuaian ini disebabkan data validasi hanya berisi 35 perusahaan bangkrut sehingga *permutation importance* untuk sebagian besar fitur bernilai mendekati nol dan tidak stabil. Meskipun demikian, *Total debt/Total net worth* dan *Persistent EPS in the Last Four Seasons* tetap menempati dua posisi teratas *permutation importance* pada RF + SMOTE dan bernilai tinggi pada RF. Temuan ini menegaskan bahwa stabilitas peringkat MDI tidak otomatis berarti stabilitas kontribusi prediktif, sehingga interpretasi sebaiknya hanya difokuskan pada rasio yang konsisten lintas metode dan skenario.

4.5 Optimasi Ambang Batas Klasifikasi (Threshold Tuning)

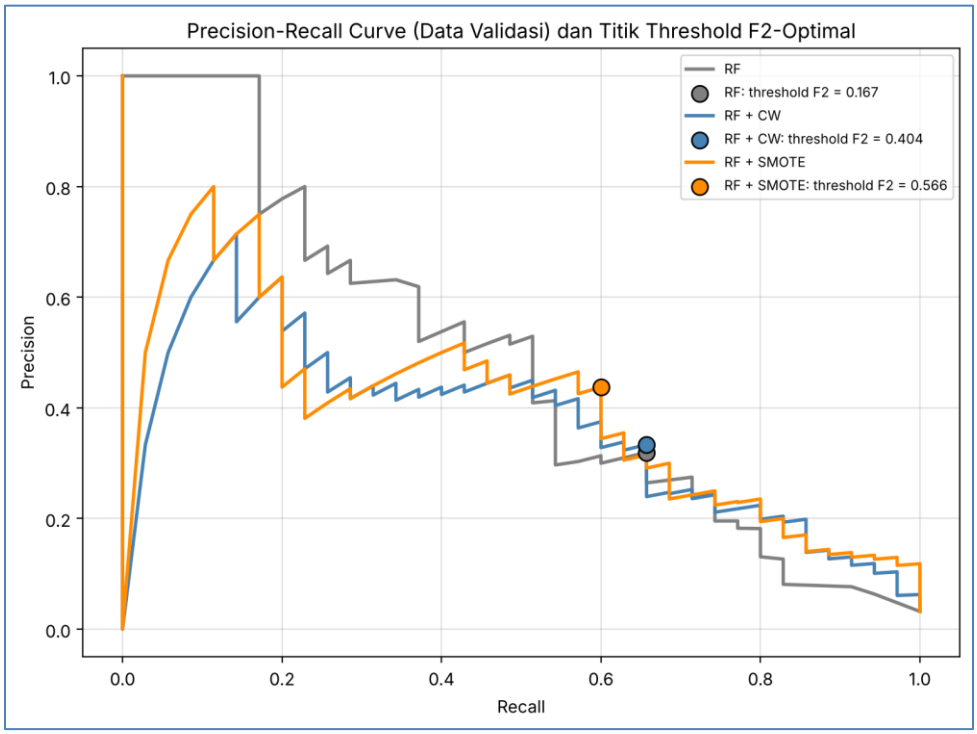
Secara bawaan, *Random Forest* mengklasifikasikan observasi sebagai kelas positif apabila probabilitas prediksinya minimal 0,5. Karena kegagalan mendeteksi perusahaan bangkrut (*False Negative*) umumnya lebih merugikan daripada *False Positive*, *threshold* setiap skenario dioptimasi pada data validasi dengan memaksimalkan F2-Score (Tabel 4 dan Gambar 11).

 ISSN:2302-8149
e-ISSN:2540-9719

Threshold F2-optimal pada RF jauh di bawah 0,5 (0,1667), sedangkan pada RF + CW (0,4037) dan RF + SMOTE (0,5658) berada di sekitar 0,5. Hal ini konsisten dengan mekanisme penyeimbangan kelas: SMOTE dan pembobotan kelas menaikkan probabilitas prediksi kelas minoritas, sehingga *threshold* bawaan sudah mendekati titik optimal, sedangkan RF tanpa penyeimbangan memerlukan *threshold* yang jauh lebih rendah untuk mencapai *Recall* yang setara.

Tabel 4 Threshold optimal pada data validasi

Skenario	Threshold F2	Precision	Recall	F2-Score	Threshold F1	F1-Score
RF	0,1667	0,3194	0,6571	0,5425	0,3033	0,5217
RF + CW	0,4037	0,3333	0,6571	0,5502	0,5033	0,4819
RF + SMOTE	0,5658	0,4375	0,6000	0,5585	0,6024	0,5128



Gambar 11 Precision-Recall curve pada data validasi dan titik threshold f2-optimal

Karena data validasi hanya berisi 35 perusahaan bangkrut, perubahan satu kasus terdeteksi mengubah *Recall* sekitar 0,029. Kurva PR pada Gambar 11 juga bergerigi, sehingga *threshold* hasil *tuning* berpotensi tidak stabil ketika diterapkan pada data uji. Oleh karena itu, evaluasi akhir melaporkan kedua kondisi, yaitu *threshold* 0,5 dan *threshold* hasil *tuning*.

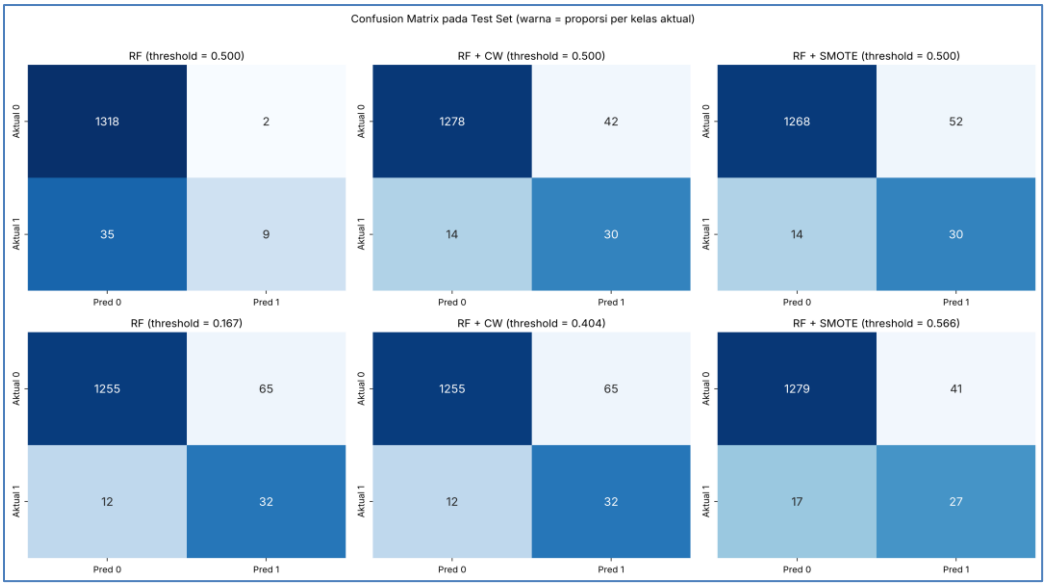
4.6 Evaluasi Akhir pada Data Uji

Evaluasi akhir dilakukan pada data uji yang tidak pernah digunakan dalam pelatihan, pemilihan *hyperparameter*, maupun *threshold tuning*. Hasil enam kondisi pengujian disajikan pada Tabel 5, Gambar 12, dan Gambar 13.

Tabel 5 Performa pada data uji (44 bangkrut, 1.320 tidak bangkrut)

Skenario	Threshold	Nilai	Precision	Recall	F1	F2	ROC-AUC	PR-AUC	FN	FP
RF	0,5	0,5000	0,8182	0,2045	0,3273	0,2406	0,9439	0,5157	35	2
RF	F2-tuned	0,1667	0,3299	0,7273	0,4539	0,5861	0,9439	0,5157	12	65
RF + CW	0,5	0,5000	0,4167	0,6818	0,5172	0,6048	0,9526	0,5347	14	42
RF + CW	F2-tuned	0,4037	0,3299	0,7273	0,4539	0,5861	0,9526	0,5347	12	65
RF + SMOTE	0,5	0,5000	0,3659	0,6818	0,4762	0,5814	0,9551	0,5024	14	52
RF + SMOTE	F2-tuned	0,5658	0,3971	0,6136	0,4821	0,5533	0,9551	0,5024	17	41

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 12 Confusion matrix enam kondisi pengujian pada data uji

Dampak SMOTE (RF vs RF + SMOTE pada threshold 0,5). RF tanpa penyeimbangan hanya mendeteksi 9 dari 44 perusahaan bangkrut (*Recall* 0,2045; 35 *False Negative*), meskipun *Precision*-nya tinggi (0,8182). Penerapan SMOTE meningkatkan *Recall* menjadi 0,6818 (14 *False Negative*), dengan selisih 0,4773 (CI 95%: 0,3409 s.d. 0,6364) yang signifikan; F2-Score juga meningkat signifikan. Peningkatan ini disertai penurunan *Precision* yang signifikan menjadi 0,3659 karena *False Positive* bertambah dari 2 menjadi 52. Uji McNemar pada kelas bangkrut menunjukkan SMOTE mendeteksi 21 perusahaan bangkrut yang tidak terdeteksi RF, tanpa satu pun kasus sebaliknya ($p < 0,001$).

Tabel 6 Paired bootstrap (2.000 replikasi): selisih metrik dan CI 95% pada data uji

Perbandingan	Metrik	Selisih	CI 95%	Signifikan
RF + SMOTE − RF (thr 0,5)	Recall	0,4773	[0,3409; 0,6364]	Ya
RF + SMOTE − RF (thr 0,5)	Precision	-0,4523	[-0,6628; -0,1923]	Ya
RF + SMOTE − RF (thr 0,5)	F2	0,3408	[0,2008; 0,4869]	Ya
RF + SMOTE − RF (thr 0,5)	ROC-AUC	0,0112	[-0,0046; 0,0307]	Tidak
RF + SMOTE − RF (thr 0,5)	PR-AUC	-0,0133	[-0,1375; 0,1211]	Tidak
RF + SMOTE − RF + CW (thr 0,5)	Recall	0,0000	[-0,0682; 0,0682]	Tidak
RF + SMOTE − RF + CW (thr 0,5)	Precision	-0,0508	[-0,1101; 0,0005]	Tidak
RF + SMOTE − RF + CW (thr 0,5)	F2	-0,0234	[-0,0806; 0,0298]	Tidak
RF + SMOTE − RF + CW (thr 0,5)	ROC-AUC	0,0025	[-0,0061; 0,0138]	Tidak
RF + SMOTE − RF + CW (thr 0,5)	PR-AUC	-0,0323	[-0,1183; 0,0489]	Tidak
RF + SMOTE: F2-tuned − 0,5	Recall	-0,0682	[-0,1364; 0,0000]	Tidak
RF + SMOTE: F2-tuned − 0,5	Precision	0,0312	[-0,0113; 0,0782]	Tidak
RF + SMOTE: F2-tuned − 0,5	F2	-0,0281	[-0,0913; 0,0251]	Tidak

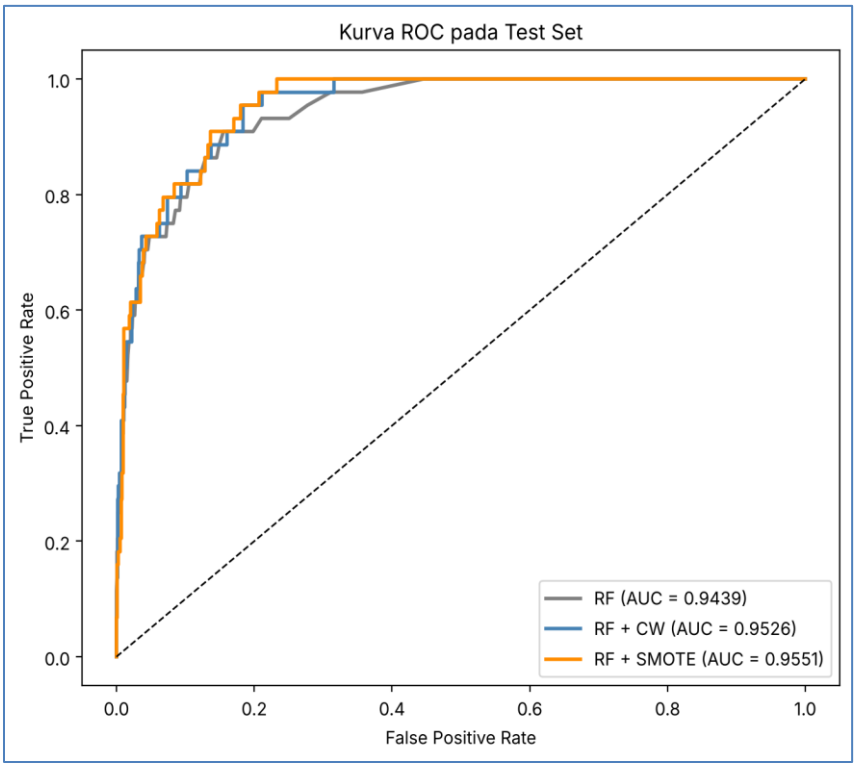
Catatan: ROC-AUC dan PR-AUC tidak bergantung pada threshold, sehingga selisihnya untuk perbandingan threshold dalam satu skenario selalu nol dan tidak ditampilkan. Uji McNemar exact: RF + SMOTE vs RF p = < 0,001 (seluruh data) dan p < 0,001 (kelas bangkrut); RF + SMOTE vs RF + CW p = 0,064 dan p = 1,000; RF + SMOTE F2-tuned vs 0,5 p = 0,057 dan p = 0,250.

Namun, ROC-AUC hanya berubah dari 0,9439 menjadi 0,9551 dan PR-AUC dari 0,5157 menjadi 0,5024; keduanya tidak signifikan karena CI memuat nol (Tabel 6). Artinya, SMOTE tidak secara nyata meningkatkan kemampuan model dalam mengurutkan perusahaan berdasarkan risiko, melainkan menggeser probabilitas prediksi kelas minoritas ke atas sehingga lebih banyak perusahaan melewati *threshold* 0,5. Mekanismenya dapat dijelaskan sebagai berikut: setelah SMOTE, proporsi kelas bangkrut pada data latih naik dari 3,2% menjadi 50%, sehingga proporsi kelas bangkrut pada daun-daun pohon dan probabilitas yang dihasilkan *Random Forest* ikut meningkat. Penjelasan ini didukung oleh RF tanpa penyeimbangan yang cukup diturunkan *threshold*-nya ke 0,1667: *Recall*-nya menjadi 0,7273 dengan F2-Score 0,5861, setara dengan RF + SMOTE (0,5814). Temuan ini sejalan dengan [24] yang menunjukkan bahwa koreksi ketidakseimbangan kelas pada dasarnya menggeser probabilitas prediksi.

 ISSN:2302-8149
e-ISSN:2540-9719

SMOTE vs pembobotan kelas (RF + CW vs RF + SMOTE). Pada *threshold* 0,5, kedua skenario mendeteksi jumlah kasus bangkrut yang sama (*Recall* 0,6818), sedangkan RF + CW menghasilkan *False Positive* lebih sedikit (42 berbanding 52). Seluruh selisih metrik antara kedua skenario tidak signifikan (Tabel 6), dan uji McNemar juga tidak menunjukkan perbedaan. Dengan demikian, pada dataset ini SMOTE tidak memberikan keunggulan yang terbukti secara statistik dibandingkan pembobotan kelas yang lebih sederhana dan tidak menambah data sintetis.

Dampak threshold tuning. Pada RF + SMOTE, *threshold* hasil *tuning* (0,5658) justru menurunkan *Recall* dari 0,6818 menjadi 0,6136 (*False Negative* naik dari 14 menjadi 17) dan menurunkan F2-Score dari 0,5814 menjadi 0,5533, meskipun *False Positive* turun dari 52 menjadi 41. Perubahan tersebut tidak signifikan (Tabel 6). Dengan demikian, klaim bahwa *threshold tuning* meminimalkan *False Negative* tidak didukung oleh hasil pada RF + SMOTE. Penyebabnya adalah *threshold* F2-optimal pada data validasi berada sedikit di atas 0,5 dan ditentukan oleh sangat sedikit kasus bangkrut, sehingga tidak tergeneralisasi ke data uji. Sebaliknya, pada RF tanpa penyeimbangan, *threshold tuning* memberikan dampak terbesar (*Recall* 0,2045 menjadi 0,7273). Hasil ini menegaskan bahwa efek SMOTE dan efek *threshold tuning* merupakan dua perlakuan yang berbeda dan sebagian saling menggantikan.



Gambar 13 Kurva ROC ketiga skenario pada data uji

Kurva ROC (Gambar 13) memperlihatkan bahwa ketiga skenario memiliki kemampuan diskriminasi yang hampir sama (ROC-AUC 0,9439, 0,9526, dan 0,9551). Nilai PR-AUC yang lebih informatif untuk data tidak seimbang juga berdekatan (0,5157, 0,5347, dan 0,5024), dengan RF + SMOTE justru yang terendah. Dibandingkan penelitian terdahulu yang melaporkan peningkatan *recall* akibat SMOTE [2], [14], hasil penelitian ini konsisten pada sisi *Recall*, tetapi menambahkan bukti bahwa peningkatan tersebut tidak diikuti peningkatan kemampuan diskriminasi yang signifikan dan dapat dicapai pula melalui pembobotan kelas atau penyesuaian *threshold*.

4.7 Interpretasi dan Rekomendasi Praktis

Berdasarkan pengulangan 10 *seed* (subbagian 3.8), rasio keuangan dikelompokkan menjadi rasio yang stabil dan penting serta rasio yang sensitif terhadap SMOTE (Tabel 7).

Kategori pertama, **rasio stabil dan penting**, terdiri atas delapan rasio yang masuk Top-10 pada seluruh *seed* di kedua skenario: *Total debt/Total net worth*, *Persistent EPS in the Last Four Seasons*, *Non-industry income and expenditure/revenue*, *Net Value Growth Rate*, *Quick Ratio*, *Net*

 ISSN:2302-8149
e-ISSN:2540-9719

Value Per Share (B), *Degree of Financial Leverage (DFL)*, dan *Interest-bearing debt interest rate*. Rasio-rasio ini mewakili dimensi solvabilitas (*Total debt/Total net worth*, DFL, *Interest-bearing debt interest rate*), profitabilitas (*Persistent EPS*, *Non-industry income and expenditure/revenue*), likuiditas (*Quick Ratio*), serta nilai dan pertumbuhan ekuitas (*Net Value Per Share (B)*, *Net Value Growth Rate*). Temuan ini sejalan dengan [5] yang menempatkan *debt-to-equity ratio*, *current ratio*, dan *return on assets* sebagai fitur terpenting, karena *Total debt/Total net worth*, *Quick Ratio*, dan *Persistent EPS* merupakan proksi yang dekat dari ketiga rasio tersebut pada dataset ini, serta dengan [15] yang menunjukkan pentingnya rasio solvabilitas dan profitabilitas dalam prediksi kebangkrutan.

Tabel 7 Pengelompokan rasio keuangan berdasarkan stabilitas peringkat lintas 10 seed

Kategori	Rasio keuangan	Peringkat rata-rata RF	Peringkat rata-rata RF + SMOTE
Stabil & penting	Total debt/Total net worth	3,1	1,3
Stabil & penting	Persistent EPS in the Last Four Seasons	2,0	1,7
Stabil & penting	Non-industry income and expenditure/revenue	7,5	3,7
Stabil & penting	Net Value Growth Rate	1,0	3,8
Stabil & penting	Quick Ratio	5,1	5,2
Stabil & penting	Net Value Per Share (B)	3,9	5,3
Stabil & penting	Degree of Financial Leverage (DFL)	6,1	7,2
Stabil & penting	Interest-bearing debt interest rate	7,3	8,0
Sensitif terhadap SMOTE	Tax rate (A)	38,0 ± 0,0	16,5 ± 2,6
Sensitif terhadap SMOTE	Operating Funds to Liability	29,4 ± 3,4	9,9 ± 1,5
Sensitif terhadap SMOTE	Operating Expense Rate	16,5 ± 1,0	35,5 ± 2,0
Sensitif terhadap SMOTE	Continuous Net Profit Growth Rate	32,5 ± 2,3	14,4 ± 1,5
Sensitif terhadap SMOTE	Cash Turnover Rate	28,6 ± 2,3	11,6 ± 1,8
Sensitif terhadap SMOTE	Long-term fund suitability ratio (A)	18,7 ± 1,5	35,3 ± 1,3
Sensitif terhadap SMOTE	Total Asset Growth Rate	33,0 ± 2,3	19,2 ± 2,1
Sensitif terhadap SMOTE	Realized Sales Gross Profit Growth Rate	21,3 ± 1,6	34,0 ± 2,0
Sensitif terhadap SMOTE	Cash/Total Assets	9,1 ± 0,3	20,3 ± 3,6

Catatan: rasio stabil & penting masuk Top-10 pada 100% seed di kedua skenario. Untuk rasio sensitif ditampilkan rata-rata ± simpangan baku peringkat antar-seed.

Kategori kedua, **rasio sensitif terhadap SMOTE**, terdiri atas sembilan rasio yang peringkatnya bergeser minimal 10 posisi melebihi dua kali simpangan baku antar-seed. Sebagian naik setelah SMOTE, misalnya *Tax rate (A)*, *Operating Funds to Liability*, *Continuous Net Profit Growth Rate*, dan *Cash Turnover Rate*, sedangkan sebagian turun, misalnya *Operating Expense Rate*, *Long-term fund suitability ratio (A)*, dan *Cash/Total Assets*. Pergeseran ini dapat dijelaskan dari cara MDI dihitung: pada data seimbang, pemisahan *node* yang mampu memisahkan perusahaan bangkrut menghasilkan penurunan *impurity* yang jauh lebih besar dibandingkan pada data tidak seimbang, sehingga fitur yang membedakan kasus minoritas memperoleh bobot lebih besar. Interpolasi SMOTE juga membentuk sampel sintetis di antara kasus bangkrut yang ada, sehingga dapat memperkuat pola yang hanya didukung sedikit observasi asli. Istilah "artefak" pada penelitian ini berarti kepentingan fitur yang muncul akibat perubahan distribusi data latih, bukan bukti bahwa rasio tersebut tidak relevan secara ekonomi. Karena MDI mengukur kontribusi fitur terhadap pemisahan *node* di dalam model, hasil ini juga tidak dapat ditafsirkan sebagai hubungan kausal antara rasio keuangan dan kebangkrutan.

Secara praktis, manajemen, investor, dan kreditor disarankan memprioritaskan pemantauan rasio pada kelompok stabil karena kepentingannya tidak bergantung pada teknik penyeimbangan data. Rasio pada kelompok sensitif tetap dapat digunakan, tetapi perlu ditafsirkan dengan hati-hati dan dikonfirmasi dengan analisis keuangan lain.

4.8 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, data hanya berasal dari perusahaan Taiwan periode 1999–2009, sehingga hasil belum tentu tergeneralisasi ke negara, periode, atau kondisi ekonomi lain, termasuk pasar Indonesia. Kedua, data dibagi secara acak tanpa memperhatikan urutan waktu, sehingga belum mencerminkan skenario prediksi ke depan. Ketiga, jumlah kasus bangkrut pada data validasi (35) dan data uji (44) kecil, sehingga CI metrik cukup lebar dan *threshold* hasil *tuning* kurang stabil. Keempat, *hyperparameter* terbaik berbeda antar-skenario, sehingga perbandingan mencerminkan dampak SMOTE beserta konfigurasi optimal yang menyertainya. Kelima, seleksi fitur berbasis korelasi menghapus sejumlah rasio yang saling berkorelasi tinggi, dan

 ISSN:2302-8149
e-ISSN:2540-9719

analisis interpretasi hanya menggunakan MDI dan *permutation importance* pada satu algoritma (*Random Forest*).

5 Kesimpulan

Penelitian ini menganalisis dampak penerapan SMOTE terhadap performa prediksi dan stabilitas *feature importance Random Forest* pada prediksi *financial distress* (kebangkrutan) perusahaan Taiwan. Terkait tujuan pertama, SMOTE meningkatkan *Recall* secara signifikan dibandingkan *Random Forest* tanpa penanganan ketidakseimbangan (0,2045 menjadi 0,6818), tetapi menurunkan *Precision* secara signifikan dan tidak meningkatkan ROC-AUC maupun PR-AUC secara signifikan. SMOTE juga tidak berbeda signifikan dengan pembobotan kelas, dan efek serupa dapat diperoleh dengan menurunkan *threshold* klasifikasi. Dengan demikian, manfaat utama SMOTE adalah menggeser batas keputusan model ke arah deteksi kelas minoritas, bukan meningkatkan kemampuan diskriminasi. *Threshold tuning* berbasis F2 pada data validasi tidak meningkatkan deteksi pada model RF + SMOTE di data uji.

Terkait tujuan kedua, SMOTE mengubah peringkat *feature importance* secara moderat (Spearman 0,644), dan perubahan tersebut lebih besar daripada variasi antar-seed sehingga bukan disebabkan keacakan model. Sebagian besar perubahan juga muncul pada pembobotan kelas, yang menunjukkan bahwa sumber utamanya adalah penyeimbangan kelas. Meskipun demikian, delapan rasio, terutama *Total debt/Total net worth* dan *Persistent EPS in the Last Four Seasons*, konsisten berada pada Top-10 di seluruh pengulangan dan direkomendasikan sebagai indikator peringatan dini yang paling reliabel, sedangkan sembilan rasio yang sensitif terhadap SMOTE perlu ditafsirkan dengan hati-hati.

Penelitian selanjutnya disarankan menguji temuan ini pada dataset lintas negara dan periode, termasuk data perusahaan Indonesia, dengan pembagian data berbasis waktu; membandingkan teknik penyeimbangan lain (misalnya Borderline-SMOTE, ADASYN, dan *undersampling*) serta algoritma lain seperti XGBoost; dan menggunakan metode interpretabilitas alternatif seperti SHAP dengan jumlah kasus validasi yang lebih besar.

Referensi

[1] B. Billy Riantono and R. Andarsyah, "Analisa Performa *Algoritma Random Forest & Logistic Regression* dalam Sistem *Credit Scoring*," *Jurnal Teknologi dan Sistem Informasi Bisnis*, Vol. 6, No. 2, pp. 438–444, Apr. 2024, DOI: 10.47233/jteksis.v6i2.1308.

[2] M. S. Keya, H. Akter, Md. A. Rahman, Md. M. Rahman, M. U. Emon, and Md. S. Zulfiker, "*Comparison of Different Machine Learning Algorithms for Detecting Bankruptcy*," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Jan. 2021, pp. 705–712, DOI: 10.1109/ICICT50816.2021.9358587.

[3] M. M. H. Sizan et al., "*Bankruptcy Prediction for US Businesses: Leveraging Machine Learning for Financial Stability*," *Journal of Business and Management Studies*, Vol. 7, No. 1, pp. 1–14, Jan. 2025, DOI: 10.32996/jbms.2025.7.1.1.

[4] K. Samara and A. Shinde, "*Bankruptcy Prediction using Machine Learning and Data Preprocessing Techniques*," *Analytics*, Vol. 4, No. 3, p. 22, Sep. 2025, DOI: 10.3390/analytics4030022.

[5] D. Tanamas and I. R. Setyawan, "Peran Rasio Keuangan dalam Memprediksi *Financial Distress*," *Jurnal Manajemen Bisnis dan Kewirausahaan*, Vol. 9, No. 1, pp. 108–119, Jan. 2025, DOI: 10.24912/jmbk.v9i1.33839.

[6] K. Rian Kumaraarta, I. N. Purnama, and N. M. Satvika Iswari, "Klasifikasi *Financial Distress* pada Perusahaan Industrials yang Terdaftar di Bursa Efek Indonesia menggunakan *Algoritma Decision Tree* Periode 2022–2024," *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 10, No. 2, pp. 2472–2480, Mar. 2026, DOI: 10.36040/jati.v10i2.17427.

[7] I. Gurnani, Vincent, F. S. Tandian, and M. S. Anggreainy, "*Predicting Company Bankruptcy using Random Forest Method*," in *2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, IEEE, Sep. 2021, pp. 1–5, DOI: 10.1109/AiDAS53897.2021.9574384.

 ISSN:2302-8149
e-ISSN:2540-9719

[8] E. Soviyani, P. Cahyaningdyah, and A. Prihadyatama, "Analisis Prediksi Kebangkrutan dengan Metode *Altman Z-Score* dan *Springate* pada Perusahaan Subsektor Tekstil dan Garmen yang Terdaftar di Bursa Efek Indonesia," *Indonesian Accounting Literacy Journal*, Vol. 6, No. 1, pp. 19–31, Nov. 2025, DOI: 10.35313/ialj.v6i1.6712.

[9] K. Ulagapriya and S. Pushpa, "*A Comprehensive Study on Ensemble-based Imbalanced Data Classification Methods for Bankruptcy Data*," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Jan. 2021, pp. 800–804, DOI: 10.1109/ICICT50816.2021.9358744.

[10] M. Erkamim, S. Suswadi, M. Z. Subarkah, and E. Widarti, "Komparasi *Algoritme Random Forest* dan *XGBoosting* dalam Klasifikasi Performa UMKM," *Jurnal Sistem Informasi Bisnis*, Vol. 13, No. 2, pp. 127–134, Oct. 2023, DOI: 10.21456/vol13iss2pp127-134.

[11] F. T. Kristanti, M. Y. Febrianta, D. F. Salim, H. A. Riyadh, and B. A. H. Beshr, "*Predicting Financial Distress in Indonesian Companies using Machine Learning*," *Engineering, Technology and Applied Science Research*, Vol. 14, No. 6, pp. 17644–17649, Dec. 2024, DOI: 10.48084/etasr.8520.

[12] M. M. Al Khalifa, "*Machine Learning Frameworks for Business Decision-Making*," in *2025 IEEE International Conference on Emerging Trends in Engineering and Computing (ETECOM)*, IEEE, 2025, DOI: 10.1109/ETECOM66111.2025.11319125.

[13] A. D. Saputra and T. Handhayani, "Prediksi Kebangkrutan Perusahaan menggunakan *Decision Tree*, *Random Forest* dan *Logistic Regression*: Analisis Rasio Keuangan sebagai Indikator Rasio," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKSI)*, Vol. 13, No. 2, Jul. 2025, DOI: 10.24912/jiksi.v13i2.34303.

[14] A. Nugroho and E. Rilvani, "Penerapan Metode *Oversampling SMOTE* pada *Algoritma Random Forest* untuk Prediksi Kebangkrutan Perusahaan," *Techno.Com*, Vol. 22, No. 1, pp. 207–214, Feb. 2023, DOI: 10.33633/tc.v22i1.7527.

[15] D. Liang, C.-C. Lu, C.-F. Tsai, and G.-A. Shih, "*Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study*," *European Journal of Operational Research*, Vol. 252, No. 2, pp. 561–572, 2016, DOI: 10.1016/j.ejor.2016.01.012.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "*SMOTE: Synthetic Minority Over-Sampling Technique*," *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321–357, 2002, DOI: 10.1613/jair.953.

[17] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "*SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-Year Anniversary*," *Journal of Artificial Intelligence Research*, Vol. 61, pp. 863–905, 2018, DOI: 10.1613/jair.1.11192.

[18] L. Breiman, *"Random Forests*," *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001, DOI: 10.1023/A:1010933404324.

[19] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "*Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution*," *BMC Bioinformatics*, Vol. 8, No. 25, 2007, DOI: 10.1186/1471-2105-8-25.

[20] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araújo, and J. Santos, "*Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches*," *IEEE Computational Intelligence Magazine*, Vol. 13, No. 4, pp. 59–76, 2018, DOI: 10.1109/MCI.2018.2866730.

[21] S. Nogueira, K. Sechidis, and G. Brown, "*On the Stability of Feature Selection Algorithms*," *Journal of Machine Learning Research*, Vol. 18, No. 174, pp. 1–54, 2018.

[22] T. G. Dietterich, "*Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms*," *Neural Computation*, Vol. 10, No. 7, pp. 1895–1923, 1998, DOI: 10.1162/089976698300017197.

[23] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "*Permutation Importance: A Corrected Feature Importance Measure*," *Bioinformatics*, Vol. 26, No. 10, pp. 1340–1347, 2010, DOI: 10.1093/bioinformatics/btq134.

[24] R. van den Goorbergh, M. van Smeden, D. Timmerman, and B. Van Calster, "*The Harm of Class Imbalance Corrections for Risk Prediction Models: Illustration and Simulation using Logistic Regression*," *Journal of the American Medical Informatics Association*, Vol. 29, No. 9, pp. 1525–1534, 2022, DOI: 10.1093/jamia/ocac093.

ISSN:2302-8149
e-ISSN:2540-9719

[25] F. Barboza, H. Kimura, and E. Altman, "*Machine Learning Models and Bankruptcy Prediction*," *Expert Systems with Applications*, Vol. 83, pp. 405–417, 2017, DOI: 10.1016/j.eswa.2017.04.006.