

Pemodelan Klasifikasi Efisiensi Kalori Berbasis Data Aktivitas dan Kondisi Fisiologis Menggunakan Random Forest dan SMOTE

Aditya Pratama Werdana^{1*}

¹Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa, Bekasi, Indonesia

Dikirimkan: 07-04-2026
Diterbitkan: 13-04-2026

Keywords:

Random Forest;
SMOTE;
Machine Learning;
Multiclass Classification.

E-mail Penulis

korespondensi:

adityawerdana5@gmail.com

Abstrak. Pemodelan klasifikasi efisiensi kalori berbasis data aktivitas dan kondisi fisiologis menjadi pendekatan penting dalam analisis kebugaran berbasis data. Dataset yang digunakan dikategorikan ke dalam tiga kelas, yaitu *High Efficiency (0)*, *Low Efficiency (1)*, dan *Moderate (2)*. Permasalahan utama yang dihadapi adalah ketidakseimbangan distribusi kelas yang berpotensi menurunkan performa model. Untuk mengatasi hal tersebut, diterapkan metode *Synthetic Minority Over-sampling Technique (SMOTE)* guna menyeimbangkan jumlah data pada setiap kelas. Setelah proses *SMOTE*, masing-masing kelas memiliki distribusi yang sama, yaitu sebanyak 750.594 data. Selanjutnya, model klasifikasi dibangun menggunakan algoritma *Random Forest* yang dikenal mampu menangani data kompleks dan mengurangi *overfitting*. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 87,17%. Performa terbaik diperoleh pada kelas *Low Efficiency* dengan nilai *precision* 0,95, *recall* 0,92, dan *f1-score* 0,93. Sementara itu, kelas *Moderate* dan *High Efficiency* menunjukkan performa yang lebih rendah, terutama pada nilai *recall* masing-masing sebesar 0,25 dan 0,09. Nilai *macro average f1-score* sebesar 0,45 menunjukkan bahwa performa model antar kelas belum seimbang, sedangkan *weighted average f1-score* sebesar 0,89 mengindikasikan performa keseluruhan yang cukup baik. Secara umum, kombinasi *Random Forest* dan *SMOTE* mampu meningkatkan akurasi model, namun masih diperlukan pengembangan lebih lanjut untuk meningkatkan kinerja pada kelas minoritas.

Abstract. *Calorie efficiency classification based on activity data and physiological conditions has become an important approach in data-driven fitness analysis. The dataset used is categorized into three classes: High Efficiency (0), Low Efficiency (1), and Moderate (2). The main challenge addressed in this study is the class imbalance problem, which can negatively affect the performance of classification models. To overcome this issue, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the number of samples in each class. After applying SMOTE, each class contained an equal number of samples, totaling 750,594 instances. A classification model was then developed using the Random Forest algorithm, which is well-known for handling complex data and reducing overfitting. The evaluation results show that the model achieved an accuracy of 87.17%. The best performance was observed in the Low Efficiency class, with a precision of 0.95, recall of 0.92, and an f1-score of 0.93. In contrast, the Moderate and High Efficiency classes showed lower performance, particularly in recall values of 0.25 and 0.09, respectively. The macro-average f1-score of 0.45 indicates that the model's performance across classes is still imbalanced, while the weighted-average f1-score of 0.89 reflects good overall*

performance. In conclusion, the combination of Random Forest and SMOTE improves classification accuracy, although further improvements are needed to enhance performance on minority classes.

1. Pendahuluan

Perkembangan teknologi di bidang kesehatan dan kebugaran telah mendorong meningkatnya pemanfaatan data aktivitas dan kondisi fisiologis dalam memahami kinerja tubuh manusia. Data seperti jumlah langkah, detak jantung, durasi aktivitas, serta parameter fisiologis lainnya kini dapat diperoleh dengan mudah melalui perangkat wearable [1]. Hal ini membuka peluang besar untuk melakukan analisis berbasis data, khususnya dalam mengukur efisiensi penggunaan energi atau kalori selama aktivitas fisik.

Efisiensi kalori merupakan indikator penting yang menggambarkan seberapa optimal tubuh dalam membakar energi selama melakukan aktivitas tertentu [2]. Tingkat efisiensi ini berbeda pada setiap individu, tergantung pada berbagai faktor seperti usia, jenis kelamin, berat badan, tingkat kebugaran, dan kondisi metabolisme [3]. Oleh karena itu, diperlukan suatu pendekatan sistematis untuk mengelompokkan tingkat efisiensi kalori secara akurat dan berbasis data.

Dalam konteks analisis data, klasifikasi menjadi salah satu teknik yang banyak digunakan untuk mengelompokkan data ke dalam kategori tertentu. Pada kasus efisiensi kalori, individu dapat diklasifikasikan ke dalam beberapa kategori, seperti rendah, sedang, dan tinggi. Klasifikasi ini tidak hanya membantu dalam pemahaman kondisi individu, tetapi juga dapat digunakan sebagai dasar dalam memberikan rekomendasi aktivitas fisik yang lebih personal.

Seiring dengan meningkatnya kompleksitas data yang dihasilkan, diperlukan metode machine learning yang mampu menangani data dengan jumlah variabel yang besar dan hubungan yang tidak linier. Salah satu metode yang banyak digunakan adalah Random Forest, yaitu algoritma berbasis ensemble learning yang menggabungkan banyak decision tree untuk meningkatkan akurasi dan stabilitas prediksi [4].

Random Forest memiliki keunggulan dalam menangani data berdimensi tinggi serta mampu mengurangi risiko overfitting dibandingkan dengan metode pohon keputusan tunggal [5]. Selain itu, metode ini juga mampu memberikan informasi mengenai pentingnya setiap fitur dalam proses klasifikasi, sehingga dapat membantu dalam memahami faktor-faktor yang paling berpengaruh terhadap efisiensi kalori.

Namun demikian, dalam praktiknya seringkali ditemukan permasalahan ketidakseimbangan data (imbalanced data), di mana jumlah data pada setiap kelas tidak seimbang [6]. Kondisi ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas, sehingga menurunkan performa klasifikasi secara keseluruhan.

Untuk mengatasi permasalahan tersebut, digunakan teknik Synthetic Minority Over-sampling Technique (SMOTE). SMOTE bekerja dengan cara menghasilkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang [7]. Dengan demikian, model yang dibangun dapat belajar secara lebih optimal dari seluruh kelas yang ada.

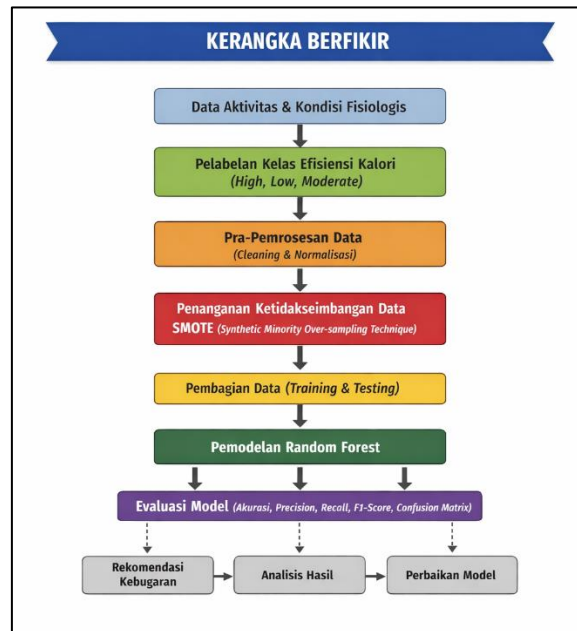
Kombinasi antara Random Forest dan SMOTE diharapkan dapat meningkatkan performa klasifikasi efisiensi kalori. Random Forest memberikan kemampuan prediksi yang kuat, sementara SMOTE membantu dalam mengatasi ketidakseimbangan data [8]. Sinergi kedua metode ini menjadi pendekatan yang efektif dalam menghasilkan model klasifikasi yang akurat dan robust.

Evaluasi model dalam penelitian ini dilakukan menggunakan berbagai metrik seperti akurasi, precision, recall, dan F1-score [9]. Penggunaan metrik yang beragam penting untuk memberikan gambaran yang lebih komprehensif mengenai performa model, terutama dalam kondisi data yang tidak seimbang.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model klasifikasi efisiensi kalori berbasis data aktivitas dan kondisi fisiologis menggunakan metode Random Forest dengan penerapan SMOTE. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem analitik kesehatan yang lebih akurat, adaptif, dan mampu mendukung pengambilan keputusan dalam bidang kebugaran dan kesehatan.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode machine learning untuk membangun model klasifikasi efisiensi kalori berdasarkan data aktivitas dan kondisi fisiologis. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data hingga evaluasi model.



Gambar 1. Kerangka Berfikir dan Desain Penelitian

2.1. Data Collection

Dataset yang digunakan merupakan data aktivitas dan kondisi fisiologis yang mencakup berbagai fitur seperti intensitas aktivitas, detak jantung, durasi, serta parameter tubuh lainnya. Data tersebut kemudian diberi label ke dalam tiga kelas efisiensi kalori, yaitu High Efficiency, Low Efficiency, dan Moderate, yang masing-masing direpresentasikan sebagai kelas 0, 1, dan 2.

2.2. Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan dalam pemodelan. Proses ini meliputi pembersihan data dari nilai kosong (missing values), normalisasi atau standarisasi jika diperlukan, serta encoding label kelas [10]. Selain itu, dilakukan analisis distribusi kelas untuk mengidentifikasi ketidakseimbangan data.

2.3. Penanganan Ketidakseimbangan Data (SMOTE)

Untuk mengatasi ketidakseimbangan kelas, digunakan metode Synthetic Minority Over-sampling Technique (SMOTE). Teknik ini bekerja dengan menghasilkan data sintesis pada kelas minoritas sehingga distribusi data menjadi seimbang. Setelah penerapan SMOTE, jumlah data pada masing-masing kelas menjadi sama [11].

2.4. Pembagian Data

Dataset yang telah diproses kemudian dibagi menjadi data latih (training set) dan data uji (testing set). Pembagian ini bertujuan untuk melatih model sekaligus menguji performanya pada data yang belum pernah dilihat sebelumnya [12].

2.5. Pemodelan Menggunakan Random Forest

Model klasifikasi dibangun menggunakan algoritma Random Forest. Metode ini merupakan teknik ensemble yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi overfitting [13]. Model dilatih menggunakan data latih yang telah diseimbangkan dengan SMOTE.

2.5. Evaluasi dan Visualisasi

Evaluasi model dilakukan menggunakan beberapa metrik, yaitu akurasi, precision, recall, dan f1-score serta digunakan confusion matrix untuk melihat distribusi prediksi pada setiap kelas [14]. Evaluasi ini bertujuan untuk mengetahui kinerja model secara menyeluruh, terutama dalam menangani setiap kelas [15].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Sedangkan visualisasi data digunakan untuk membantu memahami distribusi, pola, serta hubungan antar variabel dalam dataset, sekaligus mengevaluasi kinerja model yang dibangun [15]. Pada penelitian ini, visualisasi dilakukan melalui grafik distribusi kelas untuk melihat ketidakseimbangan data sebelum dan sesudah penerapan SMOTE, heatmap korelasi untuk mengidentifikasi hubungan antar fitur, serta confusion matrix dalam bentuk heatmap untuk mengevaluasi hasil klasifikasi [16]. Selain itu, visualisasi juga membantu dalam mengidentifikasi kelemahan model, khususnya dalam mengklasifikasikan kelas minoritas, sehingga dapat menjadi dasar dalam pengembangan model yang lebih baik [17].

3. Hasil dan Pembahasan

Pada penelitian ini, model klasifikasi efisiensi kalori dibangun menggunakan algoritma Random Forest dengan penerapan SMOTE untuk mengatasi ketidakseimbangan data. Dataset yang digunakan telah dikategorikan ke dalam tiga kelas, yaitu *High Efficiency*, *Low Efficiency*, dan *Moderate*. Sebelum penerapan SMOTE, distribusi data menunjukkan ketimpangan yang signifikan antar kelas, sehingga berpotensi menyebabkan bias pada model. Setelah dilakukan oversampling menggunakan *SMOTE*, jumlah data pada setiap kelas menjadi seimbang, masing-masing sebanyak 750.594 data.

Table 1. Dataset

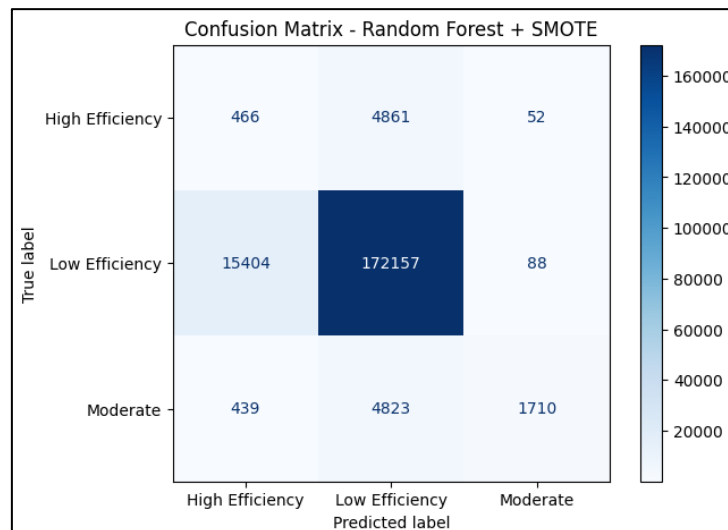
Age	Steps/Day	Calories Burned	...	Sleep Hours	Workout/Week	...	Exercise Days	Efficiency Score	Kategori
51	7853	1500	...	6.42	3	...	1	0.603	Low Efficiency
60	4820	1500	...	6.82	6	...	3	0.958	Low Efficiency
59	4251	1500	...	6.99	2	...	1	0.987	Low Efficiency
39	6275	1500	...	6.65	1	...	1	0.711	Low Efficiency
22	6490	1500	...	5.80	2	...	6	0.551	Low Efficiency
53	8517	1500	...	6.42	3	...	2	0.754	Low Efficiency
23	8263	1500	...	5.11	2	...	1	0.492	Low Efficiency
...
...
29	6901	1500	...	6.11	7	...	1	0.426	Low Efficiency

Model kemudian dilatih menggunakan data yang telah diseimbangkan dan diuji pada data uji untuk mengevaluasi performanya. Hasil pengujian menunjukkan bahwa model Random Forest mampu mencapai tingkat akurasi sebesar 87,17%. Nilai ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengklasifikasikan efisiensi kalori secara keseluruhan.

Table 2. Classification Report

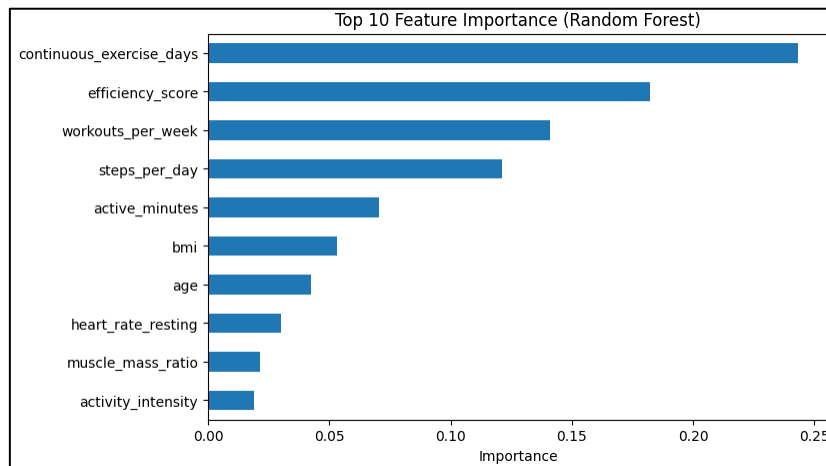
Kelas	Precision	Recall	F1-Score	Support
High Efficiency	0.03	0.09	0.04	5,379
Low Efficiency	0.95	0.92	0.93	187,649
Moderate	0.92	0.25	0.39	6,972
Accuracy			0.87	200,000
Macro Avg	0.63	0.42	0.45	200,000
Weighted Avg	0.92	0.87	0.89	200,000

Berdasarkan confusion matrix yang diperoleh, dapat dilihat bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan kelas *Low Efficiency*, dengan jumlah prediksi benar sebesar 172.157 data. Namun, model menunjukkan kecenderungan bias terhadap kelas tersebut, terlihat dari banyaknya data dari kelas lain yang salah diklasifikasikan sebagai *Low Efficiency*. Pada kelas *High Efficiency*, hanya 466 data yang berhasil diprediksi dengan benar, sementara sebagian besar lainnya (4.861 data) salah diklasifikasikan sebagai *Low Efficiency*. Hal serupa terjadi pada kelas *Moderate*, di mana hanya 1.710 data yang terklasifikasi dengan benar, sedangkan 4.823 data salah masuk ke kelas *Low Efficiency*. Kondisi ini menunjukkan bahwa meskipun model memiliki akurasi tinggi secara keseluruhan, kemampuan dalam membedakan kelas minoritas masih rendah, sehingga diperlukan perbaikan lebih lanjut untuk meningkatkan keseimbangan performa antar kelas.



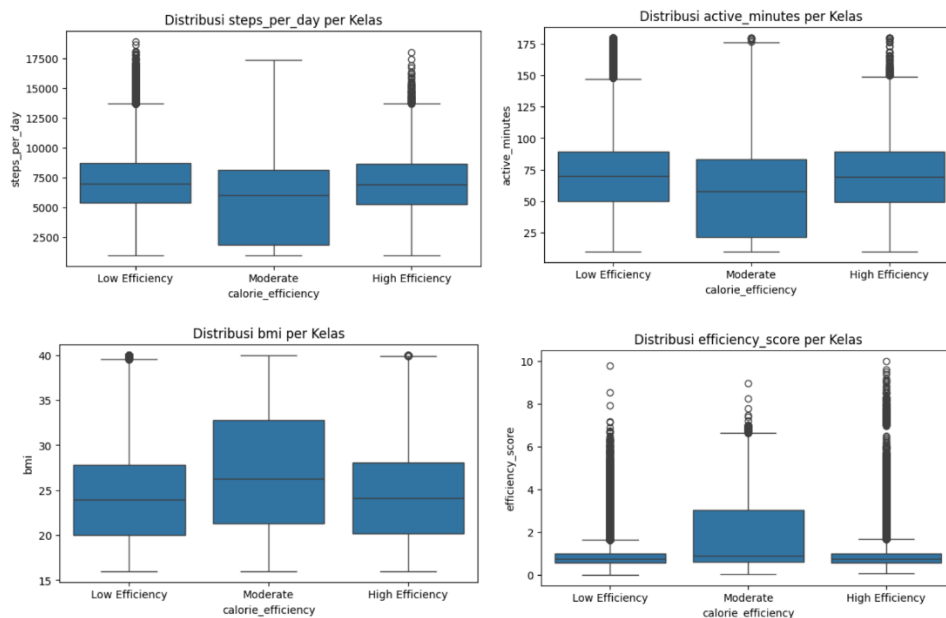
Gambar 2. Confusion Matrix

Berdasarkan visualisasi *feature importance* dari model Random Forest, dapat diketahui bahwa fitur yang paling berpengaruh dalam klasifikasi efisiensi kalori adalah *continuous_exercise_days*, diikuti oleh *efficiency_score* dan *workouts_per_week*. Hal ini menunjukkan bahwa konsistensi dalam berolahraga serta tingkat efisiensi aktivitas memiliki peran utama dalam menentukan kategori efisiensi kalori seseorang. Selain itu, fitur seperti *steps_per_day* dan *active_minutes* juga memberikan kontribusi yang cukup signifikan, yang mengindikasikan bahwa tingkat aktivitas harian turut memengaruhi hasil klasifikasi. Sementara itu, fitur seperti *bmi*, *age*, dan *heart_rate_resting* memiliki pengaruh yang lebih rendah, meskipun tetap berkontribusi dalam model. Fitur dengan kontribusi paling kecil adalah *muscle_mass_ratio* dan *activity_intensity*, yang menunjukkan bahwa variabel tersebut kurang dominan dalam menentukan efisiensi kalori pada dataset ini. Secara keseluruhan, hasil ini menunjukkan bahwa faktor kebiasaan aktivitas fisik lebih berpengaruh dibandingkan kondisi fisiologis statis dalam model yang dibangun.



Gambar 3. Feature Importance

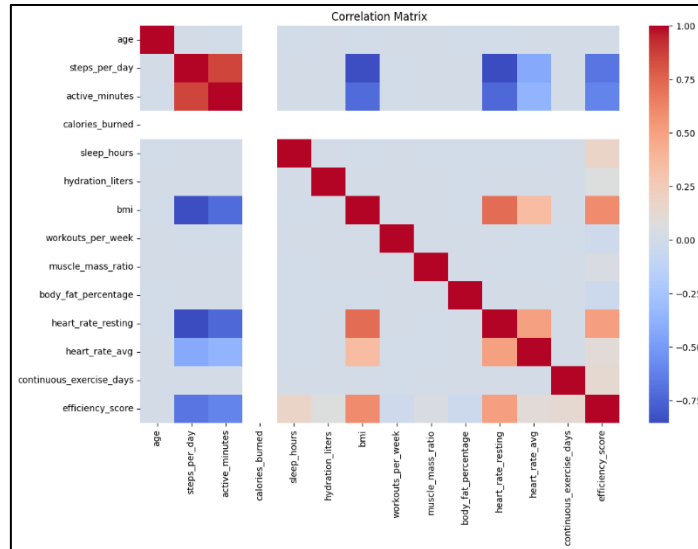
Berdasarkan visualisasi distribusi fitur, terlihat bahwa terdapat perbedaan pola antar kelas efisiensi kalori. Pada variabel *steps_per_day*, kelas *High Efficiency* dan *Low Efficiency* cenderung memiliki median yang lebih tinggi dibandingkan kelas *Moderate*, yang menunjukkan bahwa jumlah langkah harian berkontribusi terhadap peningkatan efisiensi kalori. Hal serupa juga terlihat pada *active_minutes*, di mana kelas *High Efficiency* memiliki distribusi aktivitas yang lebih tinggi, mengindikasikan bahwa durasi aktivitas berperan penting dalam efisiensi energi. Sementara itu, pada variabel *bmi*, kelas *Moderate* cenderung memiliki nilai yang lebih tinggi dan lebih bervariasi dibandingkan dua kelas lainnya, yang menunjukkan bahwa kondisi fisiologis seperti indeks massa tubuh dapat memengaruhi efisiensi kalori. Pada variabel *efficiency_score*, terlihat bahwa kelas *High Efficiency* memiliki sebaran nilai yang lebih tinggi dibandingkan kelas lainnya, meskipun terdapat banyak outlier. Secara keseluruhan, visualisasi ini menunjukkan bahwa fitur aktivitas seperti langkah harian dan durasi aktivitas memiliki pengaruh yang lebih konsisten dalam membedakan kelas dibandingkan fitur fisiologis.



Gambar 4. Visualisasi Distribusi Fitur

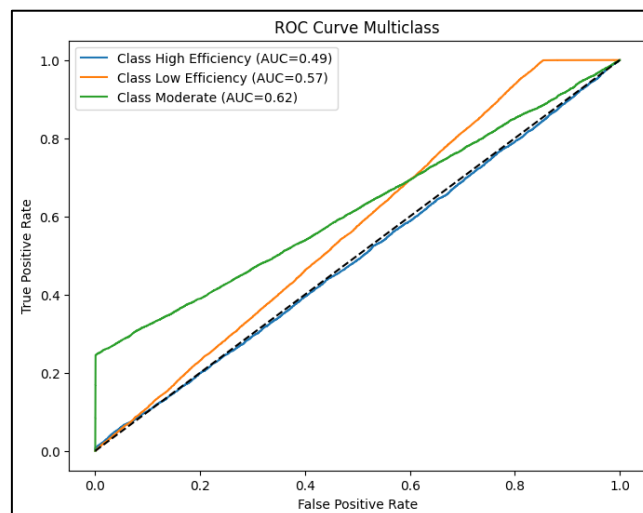
Berdasarkan visualisasi *correlation matrix*, terlihat bahwa beberapa fitur memiliki hubungan yang cukup kuat baik secara positif maupun negatif. Variabel *steps_per_day* dan *active_minutes* menunjukkan korelasi positif yang tinggi, yang mengindikasikan bahwa semakin banyak langkah yang dilakukan, semakin tinggi pula durasi aktivitas seseorang. Selain itu, kedua variabel tersebut memiliki korelasi negatif terhadap beberapa fitur lain seperti *bmi* dan *heart_rate_resting*, yang menunjukkan bahwa individu dengan aktivitas tinggi cenderung memiliki kondisi fisiologis yang lebih baik. Variabel *bmi* juga memiliki korelasi positif dengan *heart_rate_resting* dan *efficiency_score*, yang menandakan adanya hubungan antara kondisi tubuh dengan efisiensi energi. Sementara itu,

fitur seperti *sleep_hours*, *hydration_liters*, dan *muscle_mass_ratio* menunjukkan korelasi yang relatif rendah terhadap variabel lain, sehingga kontribusinya dalam model cenderung lebih kecil. Secara keseluruhan, visualisasi ini menunjukkan bahwa fitur aktivitas memiliki hubungan yang lebih kuat dibandingkan beberapa fitur fisiologis lainnya dalam menentukan efisiensi kalori.



Gambar 5. Correlation Matrix

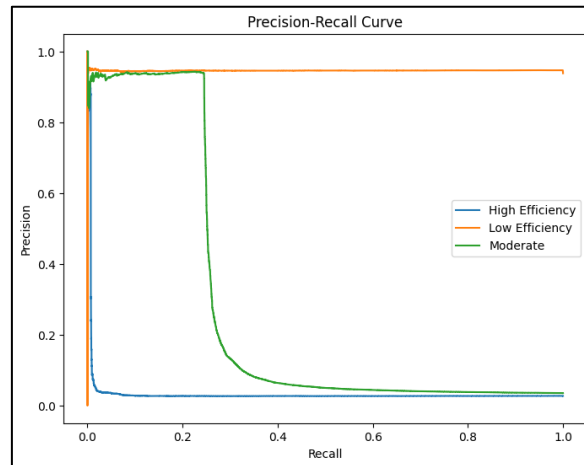
Berdasarkan visualisasi ROC Curve multiclass, terlihat bahwa performa model dalam membedakan masing-masing kelas masih relatif terbatas. Nilai AUC untuk kelas *High Efficiency* sebesar 0,49 menunjukkan bahwa model hampir tidak lebih baik dari tebakan acak dalam mengklasifikasikan kelas tersebut. Sementara itu, kelas *Low Efficiency* memiliki nilai AUC sebesar 0,57 dan kelas *Moderate* sebesar 0,62, yang mengindikasikan kemampuan klasifikasi yang sedikit lebih baik, namun masih tergolong rendah. Kurva ROC yang cenderung mendekati garis diagonal memperkuat bahwa model belum mampu memisahkan kelas secara optimal. Hal ini sejalan dengan hasil sebelumnya yang menunjukkan bahwa model masih mengalami kesulitan dalam mengenali kelas minoritas. Oleh karena itu, diperlukan perbaikan lebih lanjut, seperti tuning parameter, penambahan fitur, atau penggunaan algoritma lain, untuk meningkatkan kemampuan diskriminasi model terhadap seluruh kelas.



Gambar 6. ROC Curve Multiclass

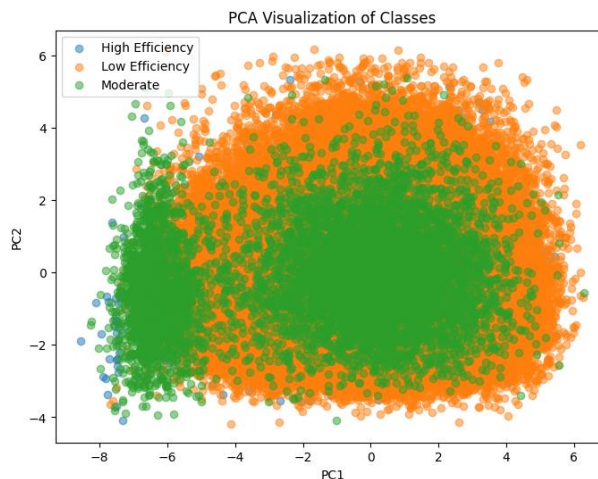
Berdasarkan visualisasi *Precision-Recall Curve*, terlihat bahwa performa model berbeda secara signifikan pada setiap kelas. Kelas *Low Efficiency* menunjukkan performa terbaik dengan nilai precision yang stabil tinggi (mendekati 0,95) pada hampir seluruh rentang recall, yang menandakan bahwa model sangat baik dalam mengidentifikasi kelas ini. Sebaliknya, kelas *High Efficiency* memiliki nilai precision yang sangat rendah dan cepat menurun, menunjukkan bahwa model kesulitan dalam membedakan kelas tersebut secara akurat. Sementara itu,

kelas *Moderate* menunjukkan performa yang cukup baik pada awalnya, namun mengalami penurunan precision yang signifikan seiring meningkatnya recall. Pola ini menunjukkan adanya trade-off antara precision dan recall, terutama pada kelas minoritas. Secara keseluruhan, hasil ini mengindikasikan bahwa model masih cenderung bias terhadap kelas mayoritas dan belum optimal dalam menangani ketidakseimbangan antar kelas.



Gambar 7. Precision-Recall Curve

Berdasarkan visualisasi PCA (*Principal Component Analysis*), terlihat bahwa distribusi data antar kelas masih mengalami tumpang tindih yang cukup signifikan. Kelas *Low Efficiency* (oranye) mendominasi area tengah hingga luas, sementara kelas *Moderate* (hijau) juga tersebar luas dan banyak beririsan dengan kelas lainnya. Di sisi lain, kelas *High Efficiency* (biru) cenderung terkonsentrasi pada area tertentu, namun tetap tidak terpisah secara jelas dari kelas lainnya. Kondisi ini menunjukkan bahwa pemisahan antar kelas dalam ruang fitur yang direduksi menjadi dua dimensi masih belum optimal. Tumpang tindih yang tinggi ini mengindikasikan bahwa fitur yang digunakan belum cukup kuat untuk membedakan kelas secara tegas, sehingga berdampak pada performa model yang masih kesulitan dalam mengklasifikasikan kelas minoritas secara akurat.



Gambar 8. PCA Visualization of Classes

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian, pemodelan klasifikasi efisiensi kalori menggunakan algoritma *Random Forest* dengan penerapan *SMOTE* mampu menghasilkan performa yang cukup baik dengan akurasi sebesar 87,17%. Penerapan *SMOTE* terbukti efektif dalam menyeimbangkan distribusi data, sehingga model dapat belajar dari setiap kelas secara lebih merata. Selain itu, fitur-fitur yang berkaitan dengan aktivitas fisik seperti *continuous_exercise_days*, *workouts_per_week*, dan *steps_per_day* terbukti memiliki pengaruh yang lebih dominan dibandingkan fitur fisiologis dalam menentukan efisiensi kalori. Namun demikian, hasil evaluasi menunjukkan bahwa performa model masih belum seimbang antar kelas. Model menunjukkan kinerja yang sangat

baik pada kelas *Low Efficiency*, tetapi masih mengalami kesulitan dalam mengklasifikasikan kelas *High Efficiency* dan *Moderate*. Hal ini diperkuat oleh hasil *confusion matrix*, *ROC curve*, dan *precision-recall curve* yang menunjukkan bahwa model cenderung bias terhadap kelas mayoritas. Selain itu, visualisasi *PCA* juga menunjukkan adanya tumpang tindih antar kelas, yang mengindikasikan bahwa pemisahan data belum optimal. Berdasarkan temuan tersebut, beberapa saran yang dapat diberikan untuk penelitian selanjutnya adalah melakukan optimasi parameter (*hyperparameter tuning*) pada model *Random Forest* untuk meningkatkan performa klasifikasi. Selain itu, dapat dilakukan eksplorasi metode lain seperti *Gradient Boosting* atau *XGBoost* yang memiliki kemampuan lebih baik dalam menangani data kompleks. Penggunaan teknik penyeimbangan data lain seperti *ADASYN* atau kombinasi metode juga dapat dipertimbangkan untuk meningkatkan performa pada kelas minoritas. Di sisi lain, penambahan dan seleksi fitur yang lebih relevan juga penting untuk meningkatkan kemampuan model dalam membedakan antar kelas. Dengan pengembangan lebih lanjut, diharapkan model yang dihasilkan dapat memiliki performa yang lebih seimbang dan akurat dalam mengklasifikasikan efisiensi kalori.

Daftar Rujukan

- [1] R. S. Pratama, I. Ramadhan, A. Wardhana, and T. N. Abdillah, "Analisis Literatur tentang Penggunaan Wearable IoT (Smart Watch, Smart Clothes, Smart Ring) untuk Optimalisasi Aktivitas Olahraga," *J. Penelit. Inov.*, vol. 5, no. 2, pp. 1857–1864, Jun. 2025, doi: 10.54082/jupin.1551.
- [2] W. H. Al Qindy, J. Y. Damaira, and S. H. L. Siregar, "Analisis langkah jalan bulanan dan kalori yang dihasilkan pada pemahaman konsep energi dan metabolisme individu," *Tak. J. Pendidikan, Sains, dan Hum.*, vol. 4, no. 2, pp. 247–256, Aug. 2025, doi: 10.56113/takuana.v4i2.143.
- [3] Y. N. A. Lestari and N. Amin, "Hubungan Status Gizi, Tingkat Kecukupan Energi dan Zat Gizi dengan Kecepatan pada Atlet Hockey Kota Surabaya," *Sport Nutr. J.*, vol. 1, no. 1, pp. 19–26, Jul. 2019, doi: 10.15294/spnj.v1i1.31275.
- [4] A. Riansah, O. Nurdiawan, and R. Herdiana, "PENERAPAN ALGORITMA RANDOM FOREST DAN DECISION TREE UNTUK MENINGKATKAN AKURASI KLASIFIKASI PENJUALAN PADA TOKO BANGUNAN," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 3, pp. 4242–4249, May 2025, doi: 10.36040/jati.v9i3.13622.
- [5] J. J. Hidayat and S. Hasanudin, "Prediksi Volatilitas IHSG Dengan Hybrid Model GARCH–Random Forest Berbasis Machine Learning," *J. Manaj. Inform. Teknol.*, vol. 6, no. 1 SE-Articles, pp. 130–140, 2026, doi: 10.51903/mifortekh.v6i1.1134.
- [6] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, p. 27, Dec. 2019, doi: 10.1186/s40537-019-0192-5.
- [7] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, Jul. 2022, doi: 10.30812/matrik.v21i3.1726.
- [8] A. Nugroho and D. Harini, "Teknik Random Forest untuk Meningkatkan Akurasi Data Tidak Seimbang," *JISITIK J. Sist. Inf. dan Teknol. Inf. Komput.*, vol. 2, no. 2, pp. 128–140, Jun. 2024, doi: 10.53624/jsitik.v2i2.379.
- [9] M. D. I. Amin, J. J. Hidayat, C. Setyowati, E. K. Fitri, A. N. Anggraini, and A. P. Werdana, "Implementasi Model LSTM Untuk Peramalan Curah Hujan Di Bekasi Dengan Pemanfaatan Data Cuaca BMKG," *J. Teknol. Inf. Digit.*, vol. 1, no. 2, pp. 90–99, [Online]. Available: <https://jurnal.ipdig.id/index.php/jtid/article/view/200>
- [10] A. H. Anshor and A. T. Zy, "Implementasi Metode Decision Tree pada Sistem Prediksi Status Kualitas Produk Minuman A," *J. Ilm. Inform. Glob.*, vol. 15, no. 1, pp. 17–22, Apr. 2024, doi: 10.36982/jiig.v15i1.3778.
- [11] N. Surojudin, S. Butsianto, and A. Firmansyah, "Perbandingan Kinerja Naïve Bayes dengan dan Tanpa SMOTE untuk Klasifikasi Gangguan Kecemasan Mahasiswa pada Data Tidak Seimbang," *Bull. Comput. Sci. Res.*, vol. 6, no. 2 SE-, pp. 804–812, Feb. 2026, doi: 10.47065/bulletincsr.v6i2.1021.
- [12] A. M. Rifai, S. Raharjo, E. Utami, and D. Ariatmanto, "Analysis for diagnosis of pneumonia symptoms using chest X-ray based on MobileNetV2 models with image enhancement using white balance and contrast limited adaptive histogram equalization (CLAHE)," *Biomed. Signal Process. Control*, vol. 90, p. 105857, Apr. 2024, doi: 10.1016/j.bspc.2023.105857.
- [13] M. Danny and M. Fatchan, "Model Prediksi Ketercapaian Learning Outcome Based Education Mahasiswa di Program Studi Teknik Informatika Menggunakan Algoritma Machine Learning," *J. Inform. Ekon. Bisnis*, pp. 684–691, Sep. 2025, doi: 10.37034/infec.v7i3.1259.
- [14] J. J. Hidayat, C. Setyowati, and A. P. Werdana, "Perancangan Sistem Prediksi Penyakit pada Tanaman Padi Berbasis Image Processing Menggunakan Algoritma VGG-16 Transfer Learning dan K-Means Segmentation," *J. Pract. Comput. Sci.*, vol. 5, no. 1, pp. 1–15, May 2025, doi: 10.37366/jpcs.v5i1.5759.
- [15] S. Arti and E. Suherlan, "Evaluasi Kinerja Machine Learning dalam Memprediksi Kemampuan Adaptasi Mahasiswa pada Lingkungan Pembelajaran Daring," *J. Pustaka AI (Pusat Akses Kaji. Teknol. Artif. Intell.)*, vol. 5, no. 1, pp. 50–57, Apr. 2025, doi: 10.55382/jurnalpustakaai.v5i1.901.
- [16] A. Bengnga and R. Ishak, "Implementasi Seleksi Fitur Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Correlation Matrix with Heatmap," *Jambura J. Electr. Electron. Eng.*, vol. 4, no. 2, pp. 169–174, Jul. 2022, doi: 10.37905/jjee.v4i2.14403.
- [17] D. P. Sari, R. Risman, F. Maulana, L. Efrizoni, and R. Rahmaddeni, "Model Prediksi Dampak Perubahan Iklim pada Ketahanan Pangan Menggunakan Algoritma Support Vector Machine and K-Nearest Neighbors," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 851–861, Jun. 2025, doi: 10.57152/malcom.v5i3.1975.