

# Rancang Bangun Aplikasi *Home Workout* dengan Prediksi Kalori Terbakar Menggunakan *Machine Learning*

Faiz Akbar

Program Studi Teknik Informatika, Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta  
Depok, Jawa Barat

[faiz.akbar.tik22@mhswn.pnj.ac.id](mailto:faiz.akbar.tik22@mhswn.pnj.ac.id)

**Abstract**— The growing use of digital technology in health and fitness has increased the need for home workout applications that not only record activity but also provide personalized analysis. However, most fitness applications still estimate burned calories using static formulas based on Basal Metabolic Rate (BMR) and Metabolic Equivalent of Task (MET), which do not adequately account for individual user characteristics. This study develops a web-based home workout application called Calsio that integrates calorie-burn prediction using a Gradient Boosting Regressor (GBR) with a fitness-nutrition chatbot based on Retrieval-Augmented Generation (RAG) using TF-IDF and the Gemini LLM in Indonesian, built with a three-tier architecture (Vue 3, Node.js/PostgreSQL, and Python-Flask). The GBR model was trained on the Gym Members Exercise Tracking dataset (973 samples), with tuning and evaluation performed entirely through K-Fold Cross Validation (K = 4, 5, 6, 8, and 10) without a separate train-test split, optimized via GridSearchCV across 96 hyperparameter combinations. The best model was obtained at learning rate 0.1 and K = 10, achieving an out-of-fold CV  $R^2$  of 0.9970, MAE of 11.62 kcal, RMSE of 14.81 kcal, and MAPE of 1.35%, clearly outperforming a Linear Regression baseline ( $R^2 = 0.9787$ , MAE = 30.21 kcal). The information retrieval system achieved a Document Retrieval Accuracy of 100% and a Mean Reciprocal Rank of 0.894. User Acceptance Testing with 10 respondents produced an average score of 4.76/5.00 (95.2%), while the System Usability Scale yielded 93.25, categorized as Grade A (Excellent). These results indicate that Calsio delivers highly accurate calorie prediction and excellent usability as an AI-based home workout solution.

**Keywords:** *Calorie Prediction; Gradient Boosting Regressor; Information Retrieval; Machine Learning; Retrieval-Augmented Generation*

**Abstrak**— Meningkatnya pemanfaatan teknologi digital di bidang kesehatan dan kebugaran mendorong kebutuhan akan aplikasi workout yang tidak hanya mencatat aktivitas, tetapi juga memberikan analisis personal. Namun, estimasi kalori pada sebagian besar aplikasi kebugaran masih menggunakan pendekatan statis berbasis rumus Basal Metabolic Rate (BMR) dan Metabolic Equivalent of Task (MET) yang belum mempertimbangkan karakteristik individual pengguna. Penelitian ini membangun aplikasi web home workout bernama Calsio yang mengintegrasikan prediksi kalori berbasis Gradient Boosting Regressor (GBR) dan chatbot fitness-nutrisi berbasis Retrieval-Augmented Generation (RAG) menggunakan TF-IDF dan LLM Gemini berbahasa Indonesia, dengan arsitektur three-tier (Vue 3, Node.js/PostgreSQL, dan Python-Flask). Model GBR dilatih pada dataset Gym Members Exercise Tracking (973 sampel), dengan tuning dan evaluasi dilakukan sepenuhnya melalui skema K-Fold Cross Validation (K=4, 5, 6, 8, dan 10) tanpa *train\_test\_split*, dioptimasi menggunakan GridSearchCV dengan 96 kombinasi parameter. Model terbaik diperoleh pada learning rate 0,1 dan K=10, dengan CV  $R^2$  (out-of-fold) sebesar 0,9970, MAE 11,62 kkal, RMSE 14,81 kkal, dan MAPE 1,35%, jauh lebih unggul dibandingkan baseline Linear Regression ( $R^2=0,9787$ , MAE=30,21 kkal). Sistem informasi retrieval mencapai Document Retrieval Accuracy 100% dan Mean Reciprocal Rank 0,894. Pengujian User Acceptance Testing terhadap 10 responden memperoleh skor rata-rata 4,76 dari 5,00 (95,2%), sedangkan System Usability Scale menghasilkan skor 93,25 (Grade A/Excellent). Hasil ini menunjukkan bahwa Calsio memiliki akurasi prediksi dan usability yang sangat baik serta layak digunakan sebagai solusi home workout berbasis kecerdasan buatan.

**Kata kunci:** *Flask; Gradient Boosting Regressor; Information Retrieval; Machine Learning; Prediksi Kalori; Retrieval-Augmented Generation; TF-IDF; Vue.js*

## I. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong transformasi digital pada berbagai sektor kehidupan, termasuk kesehatan dan kebugaran. Tingkat penetrasi internet di Indonesia mencapai 79,5% dari total populasi [1], namun Survei

Kesehatan Indonesia 2023 mencatat bahwa 37,4% penduduk usia 10 tahun ke atas masih memiliki tingkat aktivitas fisik yang kurang [2]. Kondisi tersebut mendorong popularitas *home workout* sebagai alternatif aktivitas fisik yang fleksibel dan efektif [3].

Pasar aplikasi kebugaran digital global diproyeksikan terus tumbuh hingga tahun 2030 [4]. Berbagai penelitian telah mengembangkan aplikasi kebugaran berbasis web, namun sebagian besar masih berfokus pada fungsi administratif seperti pencatatan keanggotaan dan penjadwalan, tanpa fitur analisis prediktif yang membantu pengguna memahami efektivitas latihan secara personal [5], [6].

Padahal, metode estimasi kalori yang digunakan pada sebagian besar aplikasi workout konvensional masih menggunakan pendekatan statis berbasis rumus MET dan BMR yang belum mempertimbangkan karakteristik individual pengguna, sehingga menurunkan efektivitas program workout dalam jangka panjang [7]. Penerapan *Machine Learning* menjadi pendekatan yang relevan untuk mengatasi keterbatasan tersebut; penelitian terdahulu menunjukkan bahwa model berbasis *regression* dan *ensemble learning* mampu menghasilkan performa prediksi kalori yang baik [8], [9], tetapi penelitian-penelitian tersebut masih berfokus pada pengembangan model secara terpisah dan belum diintegrasikan ke dalam aplikasi workout berbasis web yang dapat digunakan secara *real-time*.

Selain prediksi kalori, pengguna workout juga membutuhkan akses informasi mengenai gerakan, kelompok otot, dan nutrisi secara cepat. Pendekatan *Retrieval-Augmented Generation* (RAG) terbukti mampu meningkatkan relevansi dan akurasi jawaban *chatbot* dibandingkan pendekatan *Large Language Model* konvensional [10], [11]. Berdasarkan uraian tersebut, masih terdapat kesenjangan pada aplikasi workout berbasis web yang tersedia saat ini, khususnya pada aspek integrasi prediksi kalori adaptif berbasis *Machine Learning* dan penyediaan informasi workout berbasis kecerdasan buatan dalam satu sistem yang utuh.

## B. Posisi Penelitian dan Kebaruan

Kajian terhadap penelitian terdahulu menunjukkan bahwa studi prediksi kalori berbasis *Machine Learning* umumnya belum diintegrasikan ke dalam aplikasi workout harian yang utuh, sebagaimana dirangkum pada Tabel I.

| Peneliti         | Metode                        | Kekurangan vs. Penelitian Ini        |
|------------------|-------------------------------|--------------------------------------|
| Nipas et al. [8] | Random Forest Regression      | Tanpa implementasi aplikasi web      |
| Kadam et al. [9] | Random Forest Regressor       | Tanpa pencatatan workout terstruktur |
| Nasb et al. [20] | Tinjauan literatur AI-nutrisi | Konseptual, tanpa aplikasi           |
| Wicaksana [21]   | Perhitungan BMR statis        | Tanpa <i>Machine Learning</i>        |

|                        |                         |                       |
|------------------------|-------------------------|-----------------------|
| Mauliyanda et al. [22] | KNN rekomendasi workout | Tanpa prediksi kalori |
|------------------------|-------------------------|-----------------------|

Tabel I. Perbandingan dengan Penelitian Terdahulu

Nipas *et al.* [8] membandingkan Linear Regression, Ridge Regression, dan Random Forest Regression, dengan akurasi terbaik 95,77%, namun hanya berfokus pada pemilihan algoritma tanpa implementasi aplikasi. Kadam *et al.* [9] menggunakan Random Forest Regressor pada dataset lebih dari 15.000 data dengan RMSE 8,3, tetapi tanpa integrasi pencatatan workout harian secara terstruktur. Nasb *et al.* [20] bersifat konseptual dan tinjauan literatur tanpa pengembangan aplikasi, sedangkan Wicaksana [21] serta Mauliyanda *et al.* [22] masing-masing mengembangkan aplikasi workout berbasis perhitungan BMR statis dan rekomendasi berbasis KNN, tanpa model prediksi kalori adaptif. Penelitian ini mengisi kesenjangan tersebut dengan mengintegrasikan model GBR yang dioptimasi secara sistematis ke dalam aplikasi web workout harian yang utuh dan dapat digunakan pengguna secara *real-time*.

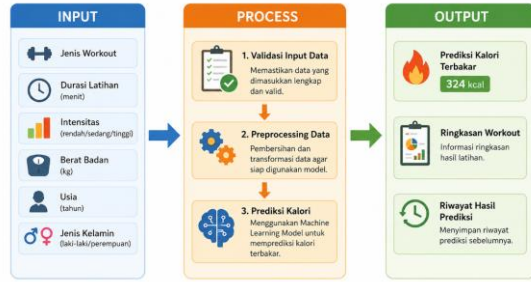
Oleh karena itu, penelitian ini merancang dan membangun aplikasi web home workout bernama Calsio, yang menggabungkan algoritma *Gradient Boosting Regressor* (GBR) untuk prediksi kalori terbakar dan sistem RAG berbasis TF-IDF serta LLM Gemini untuk fitur *chatbot* Buku Acuan berbahasa Indonesia. Kontribusi penelitian ini adalah: (1) penerapan GBR yang dioptimasi menggunakan GridSearchCV dan dievaluasi murni melalui skema K-Fold Cross Validation tanpa pemisahan data uji terpisah, (2) integrasi model prediksi ke dalam aplikasi web tiga lapisan yang dapat digunakan secara *real-time*, dan (3) evaluasi menyeluruh terhadap akurasi model, sistem information retrieval, serta usability aplikasi.

## II. METODOLOGI PENELITIAN

### A. Rancangan Penelitian

Penelitian ini menggunakan pendekatan *System Engineering* yang dikombinasikan dengan pendekatan kuantitatif eksperimental, mengikuti model pengembangan *Waterfall* yang dimodifikasi dengan siklus hidup data (*Data Science Lifecycle*). Sembilan tahapan penelitian—analisis kebutuhan, pengumpulan data, pembersihan data, pembagian dataset, pelatihan model, perancangan aplikasi, implementasi web, pengujian web, dan penulisan laporan—dipetakan terhadap enam fase CRISP-DM [12] untuk memastikan proses pengembangan model dapat dipertanggungjawabkan secara metodologis.

Objek penelitian ini adalah aplikasi workout harian dengan prediksi kalori, dengan arsitektur digambarkan melalui diagram *Input-Process-Output* (IPO) pada Gambar 1. Variabel input terdiri atas data aktivitas (jenis workout, durasi, intensitas) dan data fisiologis (berat badan, usia, jenis kelamin). Proses ini memanfaatkan algoritma Gradient Boosting Regressor untuk mengkalkulasi prediksi berdasarkan pola data historis, sedangkan output berupa nilai estimasi kalori terbakar (kcal) dan ringkasan workout yang tersimpan pada basis data.



Gambar 1. Arsitektur Input-Process-Output Aplikasi Calsio

## B. Sumber Data

Dataset yang digunakan adalah *Gym Members Exercise Tracking* yang dipublikasikan di platform Kaggle [13], terdiri dari 973 sampel dengan 15 atribut dan dinyatakan bersih dari *missing values* maupun duplikat. Atribut yang digunakan mencakup data fisiologis (usia, jenis kelamin, berat badan, tinggi badan, BMI, detak jantung, persentase lemak), data aktivitas (tipe workout, durasi sesi, frekuensi, tingkat pengalaman, konsumsi air), serta variabel target *Calories\_Burned*.

## C. Praproses dan Pembagian Data

Data kategorikal (Gender dan Workout Type) ditransformasi menjadi bentuk numerik melalui *encoding*. Normalisasi skala fitur tidak diperlukan karena algoritma *tree-based* seperti GBR bersifat *invariant* terhadap skala fitur [14]. Berbeda dengan skema konvensional yang memisahkan data latih dan data uji secara statis, penelitian ini menggunakan skema K-Fold Cross Validation (K=4, 5, 6, 8, dan 10) sehingga seluruh sampel data digunakan baik sebagai data latih maupun data validasi secara bergantian (evaluasi *out-of-fold*).

## D. Optimasi Hyperparameter

Optimasi hyperparameter dilakukan menggunakan *GridSearchCV* dari pustaka *scikit-learn* [15], yang menguji seluruh kombinasi parameter pada Tabel II dan memilih kombinasi terbaik berdasarkan proses *cross validation*. Kombinasi tersebut menghasilkan 96 skenario parameter untuk setiap konfigurasi pengujian. Penelitian ini menguji dua nilai *learning rate* (0,01

dan 0,1) pada lima variasi K-Fold (4, 5, 6, 8, dan 10), sehingga menghasilkan 10 skenario pengujian menyeluruh.

| Parameter         | Nilai yang Diuji |
|-------------------|------------------|
| n_estimators      | 100, 200, 300    |
| learning_rate     | 0,01 dan 0,1     |
| max_depth         | 2, 3, 4, dan 5   |
| min_samples_split | 2 dan 5          |
| min_samples_leaf  | 1 dan 2          |
| subsample         | 0,8 dan 1,0      |

Tabel II. Parameter GridSearchCV

## E. Metrik Evaluasi

Performa model dievaluasi menggunakan empat metrik regresi, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) [16], dan koefisien determinasi ( $R^2$ ) [17], yang secara berturut-turut dirumuskan sebagai berikut:

$$MAE = (1/n) \sum |y_i - \hat{y}_i| \quad (1)$$

$$RMSE = \sqrt{(1/n) \sum (y_i - \hat{y}_i)^2} \quad (2)$$

$$MAPE = (100\%/n) \sum |(y_i - \hat{y}_i)/y_i| \quad (3)$$

$$R^2 = 1 - (SS_{res} / SS_{tot}) \quad (4)$$

dengan  $y_i$  sebagai nilai aktual,  $\hat{y}_i$  sebagai nilai prediksi, dan  $n$  sebagai jumlah sampel. Selain metrik regresi, dilakukan pula pengujian *Black Box Testing* terhadap endpoint API menggunakan Postman, *Integration Testing*, pengujian frontend menggunakan Cypress, evaluasi sistem *Information Retrieval* (IR), serta pengujian *User Acceptance Testing* (UAT) dan *System Usability Scale* (SUS) untuk menilai kualitas aplikasi secara menyeluruh.

## F. Arsitektur Retrieval-Augmented Generation

Fitur *chatbot* Buku Acuan dibangun menggunakan mesin pencari hibrida yang menggabungkan *Information Retrieval* berbasis TF-IDF dan *Cosine Similarity* dengan LLM Gemini sebagai generator jawaban (*Retrieval-Augmented Generation*). Basis pengetahuan mencakup dokumen gerakan latihan, kelompok otot, dan nutrisi berbahasa Indonesia dan Inggris. Mekanisme *intent-based type boosting* diterapkan untuk meningkatkan relevansi dokumen yang diambil berdasarkan tipe kueri pengguna, sedangkan mekanisme *fallback* lokal digunakan apabila LLM tidak tersedia, sehingga sistem tetap dapat memberikan jawaban dasar tanpa bergantung sepenuhnya pada layanan eksternal.

# III. HASIL DAN PEMBAHASAN

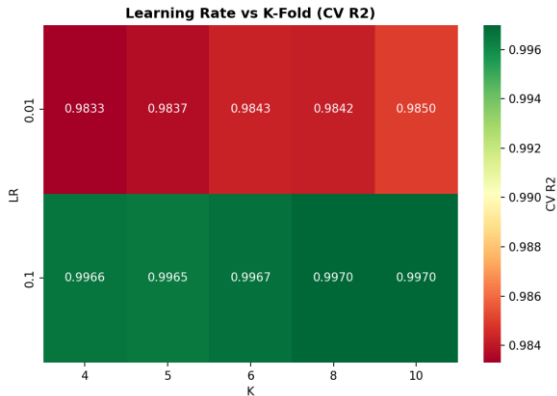
## A. Hasil Pengujian Model Gradient Boosting Regressor

Pengujian dilakukan terhadap seluruh 10 skenario kombinasi learning rate dan K-Fold Cross Validation, masing-masing dioptimasi menggunakan GridSearchCV terhadap 96 kombinasi hyperparameter. Ringkasan hasil pengujian ditampilkan pada Tabel III.

| LR   | K  | CV R <sup>2</sup> | MAE   | RMSE  | MAPE(%) |
|------|----|-------------------|-------|-------|---------|
| 0,01 | 4  | 0,9833            | 26,69 | 35,18 | 3,23    |
| 0,01 | 5  | 0,9837            | 26,17 | 34,75 | 3,17    |
| 0,01 | 6  | 0,9843            | 25,51 | 34,13 | 3,11    |
| 0,01 | 8  | 0,9842            | 25,53 | 34,25 | 3,09    |
| 0,01 | 10 | 0,9850            | 25,15 | 33,39 | 3,05    |
| 0,10 | 4  | 0,9966            | 12,02 | 15,80 | 1,42    |
| 0,10 | 5  | 0,9965            | 12,32 | 16,11 | 1,44    |
| 0,10 | 6  | 0,9967            | 12,15 | 15,69 | 1,42    |
| 0,10 | 8  | 0,9970            | 11,58 | 14,85 | 1,37    |
| 0,10 | 10 | 0,9970            | 11,62 | 14,81 | 1,35    |

Tabel III. Hasil Pengujian Seluruh Kombinasi Learning Rate dan K-Fold

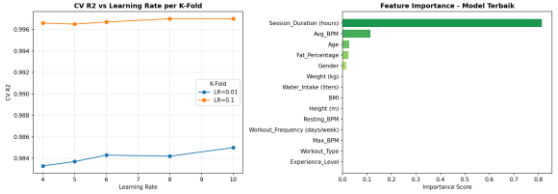
Model dengan *learning rate* 0,1 dan K=10 menghasilkan performa terbaik, yaitu CV R<sup>2</sup> (out-of-fold) sebesar 0,9970, MAE sebesar 11,62 kkal, RMSE sebesar 14,81 kkal, dan MAPE sebesar 1,35%. Learning rate 0,1 secara konsisten unggul dibandingkan 0,01 pada seluruh variasi K, sedangkan peningkatan nilai K memberikan perbaikan performa yang konsisten namun relatif kecil (selisih CV R<sup>2</sup> kurang dari 0,002 pada learning rate 0,1), mengindikasikan model yang stabil terhadap variasi jumlah fold.



Gambar 2. Heatmap CV R<sup>2</sup> pada Variasi Learning Rate dan K-Fold

Model terbaik (*learning rate* 0,1, K=10) memperoleh Train R<sup>2</sup> sebesar 0,9987 dan CV R<sup>2</sup> sebesar 0,9970, sehingga selisih (gap) hanya 0,0016—mengindikasikan model tidak mengalami overfitting maupun underfitting. Pola serupa konsisten pada seluruh skenario, dengan gap tertinggi hanya 0,0101 (*learning rate* 0,01, K=4). Sebagai pembandingan, model Linear Regression dengan skema evaluasi yang identik (K=10) hanya

menghasilkan R<sup>2</sup>=0,9787, MAE=30,21 kkal, RMSE=39,81 kkal, dan MAPE=3,76%, membuktikan bahwa GBR mampu menangkap hubungan non-linear antarfitur jauh lebih baik, dengan MAE hampir tiga kali lebih kecil.



Gambar 3. Perbandingan CV R<sup>2</sup> per Learning Rate dan Feature Importance Model Terbaik

Analisis *feature importance* pada Gambar 3 menunjukkan bahwa fitur *Session\_Duration (hours)* memberikan kontribusi paling dominan terhadap prediksi *Calories Burned*, yaitu sebesar 0,8153 (81,53%), diikuti *Avg\_BPM* sebesar 0,1149 (11,49%), *Age* sebesar 0,0276 (2,76%), *Fat\_Percentage* sebesar 0,0243 (2,43%), dan *Gender* sebesar 0,0155 (1,55%), sementara sepuluh fitur lainnya masing-masing berkontribusi kurang dari 0,11%. Temuan ini sejalan dengan hasil analisis korelasi pada tahap eksplorasi data, yang menunjukkan durasi latihan sebagai faktor paling berpengaruh terhadap kalori yang terbakar.

## B. Hasil Pengujian Aplikasi

Selain evaluasi model, dilakukan pengujian menyeluruh terhadap aplikasi Calsio yang meliputi Black Box Testing terhadap endpoint API, Integration Testing, pengujian frontend menggunakan Cypress, evaluasi sistem Information Retrieval (IR) pada fitur Buku Acuan, serta pengujian User Acceptance Testing (UAT) dan System Usability Scale (SUS). Seluruh endpoint API dan skenario integrasi berhasil memberikan respons sesuai kebutuhan aplikasi tanpa ditemukan kegagalan fungsional. Ringkasan hasil pengujian ditampilkan pada Tabel V.

| Aspek Pengujian                        | Hasil                         |
|----------------------------------------|-------------------------------|
| Black Box & Integration Testing        | Seluruh fitur berhasil (100%) |
| Document Retrieval Accuracy (IR)       | 100%                          |
| Mean Reciprocal Rank (MRR)             | 0,894                         |
| Average Precision@K                    | 63,1%                         |
| User Acceptance Testing (10 responden) | 4,76/5,00 (95,2%)             |
| System Usability Scale (SUS)           | 93,25 (Grade A)               |

Tabel V. Ringkasan Hasil Pengujian Aplikasi Calsio

Sistem IR pada fitur Buku Acuan yang menggabungkan pendekatan TF-IDF dan Cosine Similarity dengan RAG berbasis LLM Gemini mencapai Document Retrieval Accuracy sebesar

100% dan MRR sebesar 0,894 dari 26 kueri uji yang mencakup dokumen exercise, muscle, nutrisi, makro, dan MET, serta berhasil menangani variasi bahasa Indonesia-Inggris dan deteksi kueri personal. Hasil UAT dan SUS yang masing-masing berada pada kategori Sangat Baik dan Grade A (Excellent) menunjukkan bahwa integrasi prediksi kalori berbasis GBR dan chatbot berbasis RAG dapat diterima dengan baik oleh pengguna serta memenuhi tujuan penelitian dalam menyediakan solusi home workout yang akurat dan mudah digunakan.

#### IV. SIMPULAN DAN SARAN

Aplikasi web workout harian Calsio telah berhasil dikembangkan dengan mengintegrasikan dua fitur utama, yaitu prediksi kalori terbakar menggunakan model Gradient Boosting Regressor dan chatbot fitness-nutrisi berbasis Retrieval-Augmented Generation menggunakan TF-IDF dan LLM Gemini dalam Bahasa Indonesia, dibangun dengan arsitektur tiga lapisan (Vue 3, Node.js/PostgreSQL, Python-Flask). Model GBR yang dilatih pada dataset Gym Members Exercise Tracking (973 sampel) menggunakan skema K-Fold Cross Validation tanpa train\_test\_split menghasilkan performa prediksi yang sangat tinggi pada learning\_rate=0,1 dan K=10, yaitu CV R<sup>2</sup> (out-of-fold)=0,9970, MAE=11,62 kkal, RMSE=14,81 kkal, dan MAPE=1,35%, dengan gap Train R<sup>2</sup>-CV R<sup>2</sup> yang sangat kecil (0,0016) yang membuktikan model tidak mengalami overfitting.

Evaluasi menyeluruh menunjukkan performa yang sangat baik di seluruh dimensi pengujian: sistem Information Retrieval mencapai Document Retrieval Accuracy 100% dengan MRR 0,894, UAT terhadap 10 responden menghasilkan skor rata-rata 4,76/5,00 (95,2%), dan SUS menghasilkan skor 93,25 (Grade A/Excellent), membuktikan bahwa aplikasi sangat layak digunakan sebagai solusi workout harian berbasis kecerdasan buatan. Penelitian selanjutnya disarankan untuk memperluas variasi gerakan latihan, mengimplementasikan inferensi model secara on-device menggunakan TensorFlow Lite untuk meningkatkan privasi data, mengintegrasikan sinkronisasi dengan platform kesehatan seperti Google Fit atau Apple Health, serta menambahkan elemen gamifikasi untuk meningkatkan retensi pengguna.

#### V. REFERENSI

- [1] Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), "Survei Penetrasi Internet Indonesia 2024," APJII, 2024.
- [2] Kementerian Kesehatan Republik Indonesia, "Survei Kesehatan Indonesia (SKI) 2023," Badan Kebijakan Pembangunan Kesehatan, Kemenkes RI, 2024.
- [3] F. Valeriani et al., "Analysing features of home-based workout during COVID-19 pandemic: a systematic review," *Public Health*, vol. 222, pp. 100–114, 2023.
- [4] Grand View Research, "Fitness App Market Size, Share & Trends Analysis Report, 2023–2030," Grand View Research, 2023.
- [5] M. Ridwan and Z. Halim, "Perancangan Sistem Informasi Fasilitas Fitness Gym Berbasis Website Menggunakan Codeigniter," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 601–609, 2023.
- [6] S. R. Sitompul, G. P. Mahardika, W. D. A. Zebua, M. Saleh, G. M. Kombara, and S. Nugroho, "Pengembangan Sistem Manajemen Keanggotaan dan Pelatihan Gym Berbasis Web Menggunakan Metode Traditional Analysis dan Prototype," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, 2024.
- [7] A. Mahmood et al., "Development of strategies to support home-based exercise adherence after stroke: A Delphi consensus," *BMJ Open*, vol. 12, no. 1, 2022.
- [8] M. Nipas, A. G. Acoba, J. N. Mindoro, M. A. F. Malbog, J. A. B. Susa, and J. S. Gulmatico, "Burned Calories Prediction using Supervised Machine Learning: Regression Algorithm," in *Proc. ICPC2T 2022*, 2022.
- [9] A. Kadam, A. Shrivastava, S. K. Pawar, V. H. Patil, J. Michaelson, and A. Singh, "Calories Burned Prediction Using Machine Learning," in *Proc. Int. Conf. Contemporary Computing and Informatics (IC3I)*, 2023, pp. 1712–1717.
- [10] Y. Tribber, K. Kusnadi, and M. Asfi, "Implementasi Retrieval Augmented Generation untuk Layanan Informasi Kampus dengan Chatbot Berbasis AI," *Prosiding SISFOTEK*, vol. 8, no. 1, pp. 594–600, 2024.
- [11] I. Pujiono, I. M. Agtyaputra, and Y. Ruldeviyani, "Implementing Retrieval-Augmented Generation and Vector Databases for Chatbots in Public Services Agencies Context," *JITK*, vol. 10, no. 1, pp. 216–223, 2024.
- [12] P. Chapman et al., *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. Chicago, IL: SPSS Inc., 2000.
- [13] M. Rashid, "Gym Members Exercise Tracking Dataset," Kaggle, 2024.
- [14] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 507–520.
- [15] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.
- [16] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, e623, 2021.
- [17] T. O. Hodson, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE): When to use them or

- not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, 2022.
- [18] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [19] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA: O'Reilly Media, 2019.
- [20] M. Nasb, Y. Zhang, and N. Chen, "The role of artificial intelligence in precision exercise nutrition: A shift from data to diets," *Food Science and Human Wellness*, 2025.
- [21] H. Wicaksana, "Perancangan Aplikasi Workout Tanpa Alat Bantu," *Jurnal InTekSis*, vol. 10.
- [22] M. Mauliyanda, R. Deviani, F. Sabrina, and A. Afdhaluzzikri, "WorkoutLife: Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Rekomendasi Workout Berdasarkan Data Gaya Hidup," *J-SIGN*, vol. 3, no. 2, pp. 116–124, 2025.