

**EVALUASI KOMPARATIF REGRESI LINIER DAN MODEL
ENSEMBLE LEARNING UNTUK PREDIKSI KALORI
HARIAN MENGGUNAKAN DATASET GAYA HIDUP PUBLIK
DENGAN PRAPEMROSESAN TERSTRUKTUR DAN
RECURSIVE FEATURE ELIMINATION**

LAPORAN NON REGULER

SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

YUNANDRA WAHYU UTAMA

22.11.4703

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**EVALUASI KOMPARATIF REGRESI LINIER DAN MODEL
ENSEMBLE LEARNING UNTUK PREDIKSI KALORI HARIAN
MENGUNAKAN DATASET GAYA HIDUP PUBLIK DENGAN
PRAPEMROSESAN TERSTRUKTUR DAN RECURSIVE FEATURE
ELIMINATION**

LAPORAN NON-REGULER

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

YUNANDRA WAHYU UTAMA

22.11.4703

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**EVALUASI KOMPARATIF REGRESI LINIER DAN MODEL ENSEMBLE
LEARNING UNTUK PREDIKSI KALORI HARIAN MENGGUNAKAN
DATASET GAYA HIDUP PUBLIK DENGAN PRAPEMROSESAN
TERSTRUKTUR DAN RECURSIVE FEATURE ELIMINATION**

yang disusun dan diajukan oleh

Yunandra Wahyu Utama

22.11.4703

telah disetujui oleh Dosen Pembimbing:

pada tanggal 22 April 2026

Dosen Pembimbing,


Majid Rachadi, S.Kom., M.Eng
NIK. 190302393

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

JALUR NON-REGULER

EVALUASI KOMPARATIF REGRESI LINIER DAN MODEL ENSEMBLE
LEARNING UNTUK PREDIKSI KALORI HARIAN MENGGUNAKAN
DATASET GAYA HIDUP PUBLIK DENGAN PRAPEMROSESAN
TERSTRUKTUR DAN RECURSIVE FEATURE ELIMINATION

yang disusun dan diajukan oleh

Yunandra Wahyu Utama
22.11.4703

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 April 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya Dhuwita, S.Kom., M.Kom.
NIK. 190302185

Arifiyanto Hadinegoro, S.Kom., M.T.
NIK. 190302289

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 April 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, S.Kom., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Yunandra Wahyu Utama

NIM : 22.11.4703

Menyatakan bahwa Laporan dengan judul berikut:

Evaluasi Komparatif Regresi Linier dan Model Ensemble Learning Untuk Prediksi Kalori Harian Menggunakan Dataset Gaya Hidup Publik Dengan Prapemrosesan Terstruktur Dan Recursive Feature Elimination

Dosen Pembimbing : Majid Rahadi, S.Kom., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 April 2026

Yang Menyatakan,



Yunandra Wahyu Utama

HALAMAN PERSEMBAHAN

Dengan segala rasa syukur, hormat, dan penghargaan atas perjalanan panjang yang telah dilalui, laporan ini penulis persembahkan sebagai hasil dari doa, usaha, keteguhan, dan proses belajar yang tidak selalu mudah, namun penuh makna, kepada :

1. Sembah sujud serta syukur kepada Allah SWT, atas rahmat, pertolongan, dan ketetapan-Nya yang senantiasa menguatkan penulis dalam setiap tahap penyusunan laporan ini.
2. Orang tua tercinta, atas kasih sayang, doa yang tidak pernah terputus, serta dukungan tulus yang menjadi alasan penulis untuk terus melangkah.
3. Diri sendiri, atas keberanian untuk tetap berjalan, bertahan, dan menyelesaikan setiap proses hingga sampai pada titik ini.
4. Seseorang yang telah kebersamai penulis sampai saat ini, atas dukungan, perhatian, dan semangat yang diberikan.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya laporan non-reguler ini dapat diselesaikan. Laporan ini disusun sebagai salah satu bentuk pemenuhan tugas akademik dengan judul “Evaluasi Komparatif Regresi Linier dan Model Ensemble Learning Untuk Prediksi Kalori Harian Menggunakan Dataset Gaya Hidup Publik Dengan Prapemrosesan Terstruktur Dan Recursive Feature Elimination”.

Dalam proses penyusunan laporan ini, penulis memperoleh bantuan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Prof. Dr. Mohammad Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Kusriani, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Eli Pujastuti, M.Kom. selaku Ketua Program Studi Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Majid Rahardi, S.Kom., M.Kom. selaku dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan dalam penyusunan laporan ini.
5. Tim Dosen Penguji, atas kritik, saran, dan masukan yang membangun dalam penyempurnaan penelitian ini.
6. Bapak/Ibu Dosen Universitas Amikom Yogyakarta yang telah memberikan ilmu dan wawasan selama proses perkuliahan.

Yogyakarta, 10 April 2026

Penulis

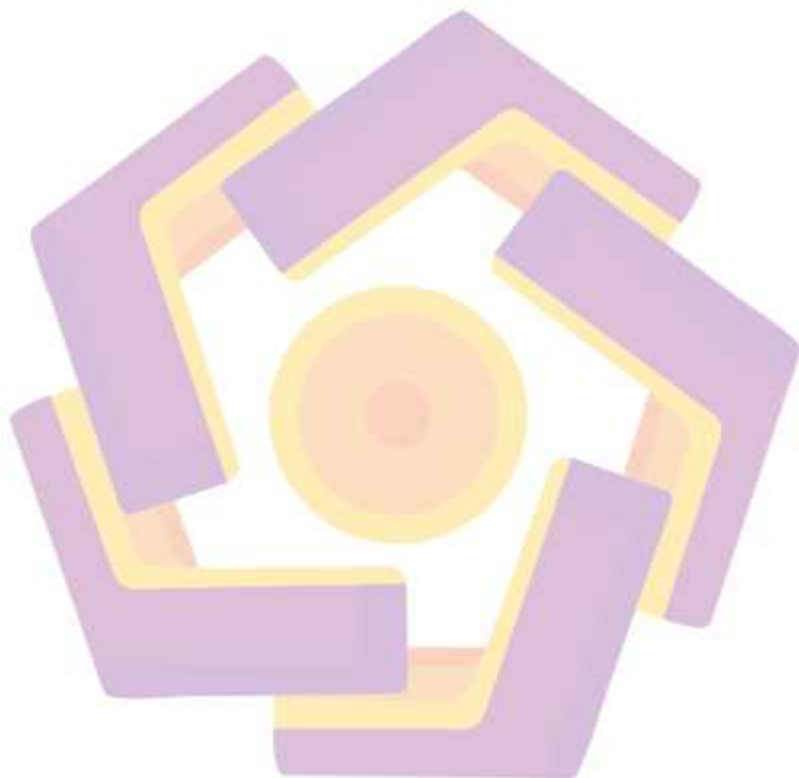
DAFTAR ISI

Halaman Judul.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN KARYA.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN.....	xii
DAFTAR ISTILAH.....	xiii
INTISARI.....	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1. Gambaran Umum.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1. Studi Literatur.....	4
2.2. Landasan Teori.....	8
2.2.1. Kalori.....	8
2.2.2. Machine Learning.....	9
2.2.3. Supervised Learning.....	9
2.2.4. Exploratory Data Analysis.....	9
2.2.5. Data Preprocessing.....	10
2.2.6. Feature Selection Recursive Feature Elimination (RFE).....	10
2.2.7. Regresi.....	10
2.2.8. Linear Regression.....	11

2.2.9. Ensemble Learning.....	11
2.2.10. Random Forest	11
2.2.11. XGBoost.....	11
2.2.12. LightGBM.....	12
2.2.13. Hyperparameter Tuning (GridSearchCV).....	12
2.2.14. Evaluasi Model.....	13
BAB III METODE PENELITIAN.....	14
3.1. Objek Penelitian.....	14
3.2. Metode Penelitian.....	14
3.2.1. Data Acquisition.....	15
3.2.2. Preprocessing Data.....	15
3.2.3. Feature Selection (RFE).....	16
3.2.4. Modeling	18
3.2.5. Evaluation	20
BAB IV HASIL DAN PEMBAHASAN.....	21
4.1. Data Acquisition.....	21
4.2. Preprocessing Data.....	21
4.2.1. Encoding Kategorikal.....	21
4.2.2. Deteksi dan Penanganan Outlier	22
4.3. RFE Feature Selection.....	24
4.4. Modeling.....	26
4.5. Evaluasi.....	27
BAB V KESIMPULAN.....	31
5.1. Kesimpulan	31
5.2. Saran.....	31
REFERENSI	33
CURICULUM VITAE	40
LAMPIRAN DAN BUKTI PENDUKUNG	41

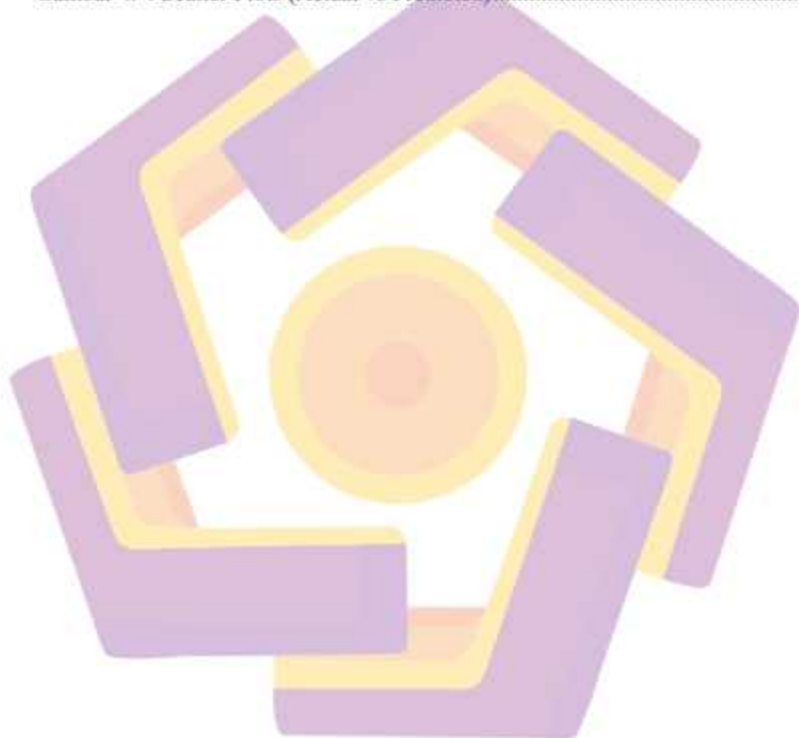
DAFTAR TABEL

Tabel 3. 1 Parameter Grid dan Hasil Best Parameters GridSearchCV	19
Tabel 4. 1 Hasil Encoding Fitur Kategorikal	21
Tabel 4. 2 Hasil Seleksi Fitur RFE.....	25
Tabel 4. 3 Hasil Evaluasi.....	27
Tabel 4. 4 Hasil Evaluasi Setelah Hyperparameter Tuning GridSearchCV	28



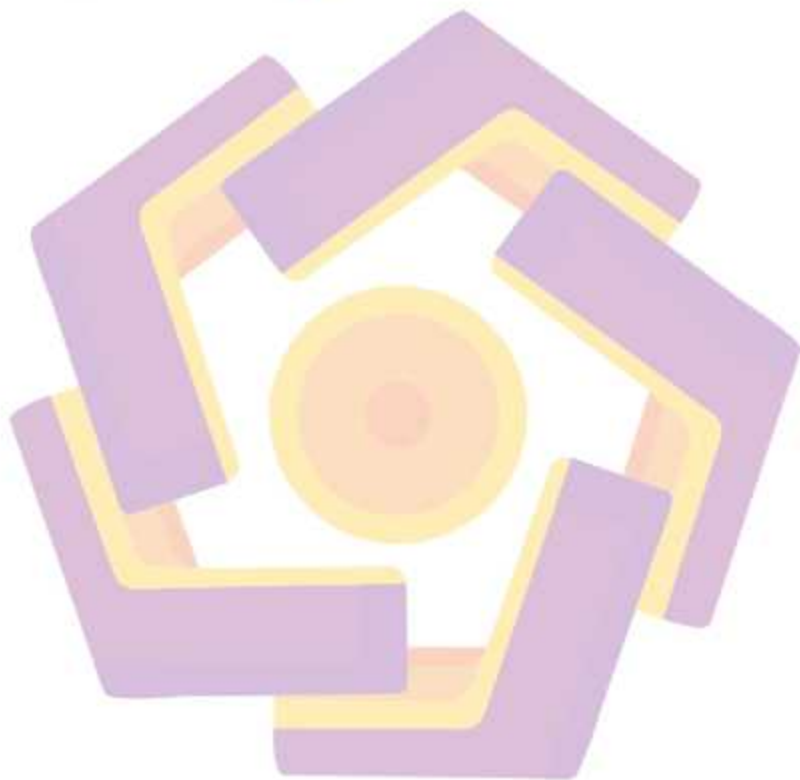
DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	15
Gambar 4. 1 Identifikasi Outlier.....	22
Gambar 4. 2 Penanganan Outlier.....	23
Gambar 4. 3 Permutation Importance.....	29
Gambar 4. 4 Scatter Plots (Actual vs Predicted).....	30



DAFTAR LAMPIRAN

Lampiran 1 Bukti LoA	41
Lampiran 2 Bukti Terindex	42
Lampiran 3 Bukti Review	47
Lampiran 4 Bukti Pembayaran.....	48



DAFTAR LAMBANG DAN SINGKATAN

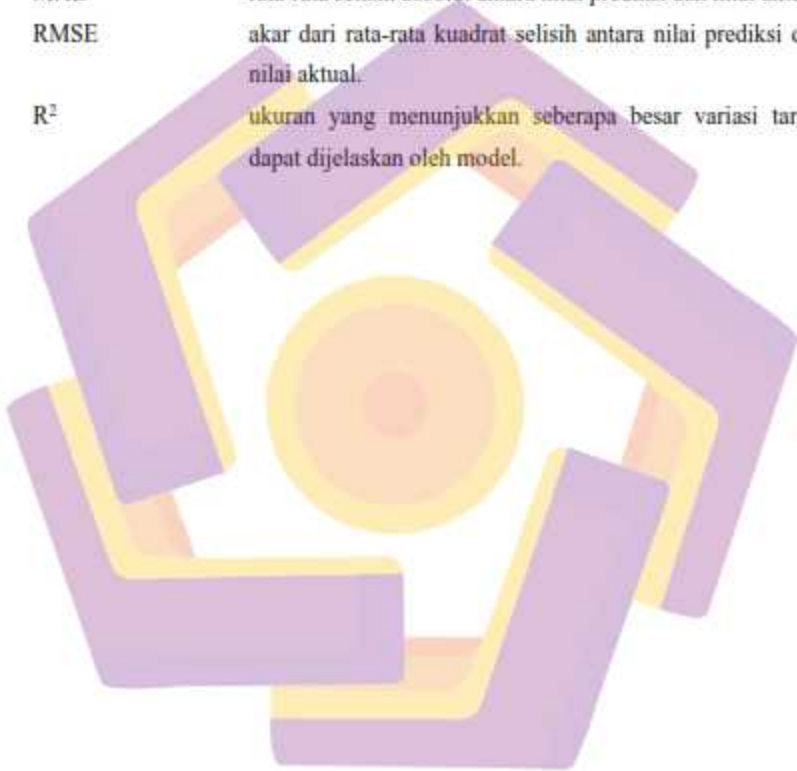


AI	<i>Artificial Intelligence</i>
BMI	<i>Body Mass Index</i>
BMR	<i>Basal Metabolic Rate</i>
CSV	<i>Comma-Separated Values</i>
EDA	<i>Exploratory Data Analysis</i>
IQR	<i>Interquartile Range</i>
LightGBM	<i>Light Gradient Boosting Machine</i>
MAE	<i>Mean Absolute Error</i>
ML	<i>Machine Learning</i>
R^2	<i>Koefisien Determinasi</i>
REE	<i>Resting Energy Expenditure</i>
RFE	<i>Recursive Feature Elimination</i>
RMSE	<i>Root Mean Squared Error</i>
XGBoost	<i>Extreme Gradient Boosting</i>

DAFTAR ISTILAH

Kalori	satuan energi yang diperoleh dari makanan dan digunakan tubuh untuk menjalankan aktivitas serta proses metabolisme.
Machine Learning	cabang kecerdasan buatan yang memungkinkan sistem mempelajari pola dari data untuk melakukan prediksi secara otomatis.
Supervised Learning	metode pembelajaran mesin yang menggunakan data berlabel sebagai dasar pelatihan model.
Preprocessing Data	tahapan penyiapan data sebelum pemodelan, seperti pembersihan, transformasi, encoding, dan normalisasi.
Encoding Kategorikal	proses mengubah data kategori menjadi bentuk numerik agar dapat diproses oleh algoritma machine learning.
Outlier	nilai data yang menyimpang jauh dari mayoritas sebaran data.
IQR capping	metode penanganan outlier dengan membatasi nilai ekstrem berdasarkan rentang interkuartil.
Normalisasi	proses penyamaan skala data agar tiap fitur memiliki rentang yang sebanding.
RFE	teknik seleksi fitur dengan menghapus fitur yang kontribusinya paling rendah secara bertahap.
Linear Regression	algoritma regresi yang memodelkan hubungan linier antara variabel input dan target.
Random Forest	algoritma ensemble berbasis banyak pohon keputusan untuk meningkatkan akurasi prediksi.
XGBoost	algoritma boosting yang membangun model secara bertahap untuk memperbaiki kesalahan prediksi sebelumnya.
LightGBM	algoritma boosting yang dirancang cepat dan efisien untuk pemodelan berbasis pohon keputusan.
Hyperparameter	

Tuning	proses mencari kombinasi parameter model terbaik untuk memperoleh performa optimal.
GridSearchCV	metode pencarian hyperparameter terbaik dengan menguji berbagai kombinasi parameter menggunakan cross-validation.
MAE	rata-rata selisih absolut antara nilai prediksi dan nilai aktual.
RMSE	akar dari rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual.
R^2	ukuran yang menunjukkan seberapa besar variasi target dapat dijelaskan oleh model.



INTISARI

Estimasi kalori harian yang akurat penting untuk mendukung personalisasi nutrisi dan pencegahan masalah kesehatan terkait pola makan. Namun, keragaman gaya hidup antarindividu membuat rekomendasi umum sering kali kurang efektif. Penelitian ini bertujuan mengembangkan model estimasi kalori berbasis gaya hidup dengan membandingkan pendekatan linear yang mudah diinterpretasikan dengan metode machine learning ensemble. Dataset publik dari Kaggle digunakan, mencakup variabel demografi, antropometri, pola makan, dan aktivitas fisik. Tahap preprocessing meliputi penanganan outlier menggunakan interquartile range capping, encoding data kategorikal, normalisasi, dan seleksi fitur menggunakan Recursive Feature Elimination (RFE). Empat model, yaitu Linear Regression, Random Forest, XGBoost, dan LightGBM, dilatih dan dievaluasi menggunakan R^2 , Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), serta waktu pelatihan. Hasil penelitian menunjukkan bahwa Linear Regression memberikan performa terbaik dengan $R^2 = 0,9650$, MAE = 80,95, RMSE = 101,71, dan waktu pelatihan 8,95 detik. Di antara model ensemble, XGBoost hasil tuning menjadi yang terbaik dengan $R^2 = 0,9646$, MAE = 81,34, RMSE = 102,35, dan waktu pelatihan 10,55 detik. Dibandingkan XGBoost, Linear Regression lebih baik dengan selisih MAE 0,39, RMSE 0,64, peningkatan R^2 sebesar 0,0004, dan waktu pelatihan lebih cepat 1,60 detik. Temuan ini menunjukkan bahwa model linear dapat menyamai bahkan melampaui model ensemble yang lebih kompleks pada data gaya hidup terstruktur.

Kata kunci: Calorie Estimation, Feature Selection, Lifestyle Data, Regression Analysis, Supervised Learning.

ABSTRACT

Accurate daily calorie estimates are essential for personalized nutrition and prevention of diet-related conditions, yet lifestyle variability can reduce the effectiveness of one-size-fits-all recommendations. This study aims to develop an accurate lifestyle-based calorie estimation model by comparing an interpretable linear approach with ensemble machine learning methods. A publicly available lifestyle dataset from Kaggle was used, containing demographic variables, anthropometric measurements, food intake, dietary patterns, and physical activity attributes. A preprocessing pipeline was applied, including outlier handling using interquartile range capping, categorical encoding, normalization, and feature selection via Recursive Feature Elimination to identify the most relevant predictors. Four models (Linear Regression, Random Forest, XGBoost, and LightGBM) were trained and evaluated, followed by hyperparameter tuning of ensemble models using GridSearchCV. Performance was assessed using R^2 , Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) and training time. Linear Regression achieved the best overall performance ($R^2 = 0.9650$, MAE = 80.95, RMSE = 101.71, training time = 8.95 seconds). Among ensembles, the tuned XGBoost performed best ($R^2 = 0.9646$, MAE = 81.34, RMSE = 102.35, training time = 10.55 seconds). Compared with tuned XGBoost, Linear Regression was superior with MAE by 0.39 and RMSE by 0.64, while R^2 increased by 0.0004 and required less computational time, indicating that added complexity did not yield meaningful gains on this structured dataset. These findings suggest that, for structured lifestyle data, interpretable linear models can match or outperform complex ensembles while remaining computationally efficient for real-time or edge-deployed health applications.

Keyword: Calorie Estimation, Feature Selection, Lifestyle Data, Regression Analysis, Supervised Learning.