Calorie Burn Prediction and Fitness Insights with Machine Learning

Sindhu S L[1], Sheethal K S[2]

[1]*Assistant Professor, Department of MCA, BIET, Davanagere*

[2]*Student, 4th Semester MCA, Department of MCA, BIET, Davanagere*

Abstract

Accurate prediction of calorie expenditure during physical activity is crucial for effective weight management and personalized fitness planning. This paper presents a machine learning- driven system that leverages an XGBoost Regressor model to estimate calories burned during exercise based on user demographics (age, gender, height, weight) and exercise-related features (duration, heart rate, body temperature). The proposed platform not only predicts calorie burn but also incorporates Body Mass Index (BMI) calculation to provide users with a holistic understanding of their health status. To further enhance user engagement, the system offers tools for estimating exercise duration required to burn a target number of calories and for projecting the time needed to achieve specific weight loss goals based on caloric deficit principles. Distinct roles are defined for Admins and Users: Admins manage user accounts, health tips, and exercise data, while Users can register, log in, access personalized calorie predictions, view health tips, and explore exercise information. Experimental results demonstrate the model's predictive accuracy and the platform's potential to support informed, data-driven decisions for fitness and weight management. This integrated approach empowers users to optimize their exercise routines and track progress, fostering improved health outcomes through personalized recommendations.

Keywords:

Calorie prediction, machine learning, XGBoost Regressor, fitness management, BMI, personalized health, caloric deficit, exercise duration.

I. INTRODUCTION

Calorie management and personalized fitness insights have become central to modern health and wellness initiatives, driven by the increasing prevalence of lifestyle-related diseases such as obesity, diabetes, and cardiovascular disorders. Accurate estimation of calories burned during physical activity is essential for effective weight management, performance optimization, and overall health improvement. Traditionally, calorie expenditure has been estimated using generalized formulas or wearable devices, but these approaches often lack the precision and personalization required to address individual variability in physiology and exercise habits.

Recent advances in artificial intelligence and machine learning have paved the way for data-

driven solutions that can analyze complex, multidimensional datasets to provide tailored health recommendations. Machine learning models, when trained on diverse user data—such as age, gender, height, weight, exercise duration, heart rate, and body temperature—can capture subtle relationships between physiological and activity-related factors, enabling more accurate and individualized predictions of calorie expenditure. This personalized approach empowers users to make informed decisions about their exercise routines, optimize their fitness plans, and track their progress with greater confidence.

Despite these technological advancements, several challenges persist in the domain of calorie burn prediction. First, the variability in user physiology—such as differences in metabolic rate, body composition, and fitness level—makes it difficult to develop a one-size-fits-all model for calorie estimation. Second, the quality and availability of reliable training data can significantly impact the performance of machine learning algorithms. Third, user engagement and system usability are critical for ensuring that individuals can effectively leverage predictive insights to achieve their health goals.

To address these challenges, we propose a comprehensive machine learning-based system for calorie burn prediction and fitness insights. The core of the system is an XGBoost Regressor model trained on a rich dataset that incorporates user demographics and real-time exercise data. By leveraging these features, the model can generate accurate, individualized predictions of calories burned during various physical activities. Beyond calorie estimation, the system integrates Body Mass Index (BMI) analysis, offering users a holistic perspective on their health status and how body composition influences exercise outcomes.

The platform is designed with two primary user roles: Admin and User. Admins are responsible for managing user accounts, curating health tips, and maintaining exercise data, ensuring the platform remains up-to-date and relevant. Users can register, log in, access personalized calorie predictions, view curated health tips, and explore a repository of exercise information. Additionally, the system provides tools for estimating the duration of exercise required to burn a target number of calories and for projecting the time needed to achieve specific weight loss goals, based on established caloric deficit principles.

II. RELATED WORK

A Study on Calories Burnt Prediction Using Machine Learning, Authors: Punita Panwar, Kanika Bhutani, Rimjhim Sharma, Rohit Saini

In today's fast-paced world, people often neglect their health and consume more junk food, leading to increased calorie intake and rising obesity rates. To address this, a machine learning system has been developed to estimate the number of calories burned during exercise using user input attributes. Trained on over 15,000 data points with a Mean Absolute Error (MAE) of 1.48, the system aims to motivate individuals by tracking their daily progress and encouraging consistent physical activity.[1]

Calorie Burn Prediction: A Machine Learning Approach using Boost XG Regression, Authors:

Niharikareddy Meenigea

With today's busy lifestyles, people often neglect physical activity and healthy eating, leading to rising obesity levels. While tracking calorie intake is relatively easy, monitoring calories burned is more challenging due to limited tools. This study aims to predict calories burned using a machine learning approach, specifically an XGBoost regression model. Trained on over 15,000 data points, the model currently achieves a Mean Absolute Error (MAE) of 2.7, which is expected to improve with more data, providing users with a reliable way to monitor their fitness progress.[2]

A Study on Calories Burnt Prediction Using Machine Learning Punita Panwar, Kanika Bhutani, Rimjhim sharma and Rohit Saini

In today's fast-paced technological world, people often overlook their health, consuming more junk food due to time constraints, leading to increased calorie intake and obesity. To address this, a machine learning system has been developed to estimate the number of calories burned after exercise using input attributes. This encourages users to stay active by tracking their progress. Trained on over 15,000 data points, the system achieves a Mean Absolute Error (MAE) of 1.48, with potential for improvement as more data is added.[3]

MACHINE LEARNING APPROACHES FOR CALORIE BURN ANALYSIS AND PREDICTION Neha
Valmiki1 , Ambili P S

This project aims to accurately predict the number of calories burned during various exercises such as running, cycling, swimming, and HIIT, which are known for their high calorie-burning potential. Given the challenge of individual differences in physiology, a dataset from the UCI Machine Learning Repository was used to develop predictive models. By applying machine learning algorithms like XGBoost Regressor and Linear Regression, the study found that these models effectively estimate calorie expenditure, outperforming other methods in accuracy.[4]

Calorie Burn Predection using Machine Learning Suvarna Shreyas Ratnakar , Vidya S

In today's fast-paced lifestyle, many people struggle to maintain healthy eating habits and regular physical activity, leading to widespread obesity. While tracking calorie intake is straightforward, monitoring calories burned is more difficult due to limited tools. This study aims to predict calories burned using an XGBoost regression model, addressing this gap with improved accuracy. Trained on over 15,000 data points, the model currently has a Mean Absolute Error (MAE) of 2.7, which is expected to improve as more data is added, offering a practical solution for fitness tracking.[5]

Calorie Burnt Prediction using XGBoost & Exercise Duration, Dasam Mohan Murali Babu

In today's busy lifestyle, people often neglect healthy eating and regular exercise, leading to rising obesity levels. While tracking calorie intake is simple, monitoring calories burned remains a challenge due to limited tools. This study addresses the issue by using an XGBoost regression machine learning model to predict calories burned based on individual exercise data. Trained on over 15,000

records, the model currently achieves a Mean Absolute Error (MAE) of 2.7, with potential for improved accuracy as more data is added.[6]

Comparing machine learning algorithms for predicting calories burned Gudiwada vijayalakshmi,Torati Sridurga, Publisher: 2023 JETIR March 2023, Volume 10, Issue 3 www.jetir.org (ISSN-2349-5162)

Calorie burn prediction plays a vital role in health and fitness by offering accurate insights for weight loss and management. Since the body burns more calories during activity and fewer in an inactive state, understanding individual energy expenditure is crucial. This system focuses on personalized factors such as age, gender, height, weight, duration, heart rate, and body temperature to ensure accurate predictions. By comparing various supervised machine learning models on the same dataset, the study identifies the most effective algorithm for predicting calories burned and highlights why others are less suitable.[7]

A Study on Calories Burnt Prediction Using Machine Learning Punita Panwar, Kanika Bhutani, Rimjhim sharma and Rohit Saini

In today's fast-paced technological era, people often neglect their health and consume unhealthy food due to time constraints, leading to increased calorie intake and obesity. To address this, a machine learning system has been developed to estimate calories burned after exercise using various input attributes. This helps users track their progress and stay motivated to maintain physical activity. Trained on over 15,000 data points, the model achieves a Mean Absolute Error (MAE) of 1.48, with accuracy expected to improve as more data is added.[8] Calories Burnt Prediction: A Machine Learning Approach Md Nahid Hosain Likhon, Farhan Bhuiyan, Md. Sihab Bhuiyan , Monjurul Aziz Fahim , and Amzad Hossain

In our research on calorie burn prediction, we applied machine learning techniques to estimate calories burned based on personal data. After performing exploratory data analysis and cleaning the dataset, we trained five models— KNN, Decision Tree, AdaBoost, SVM, and XGBoost—using both default and optimized hyperparameters. We also employed Explainable AI to understand the impact of different features on predictions. Among all models, XGBoost delivered the best performance with an RMSE of 2.13 and an R^2 score of 1, demonstrating the potential of machine learning in enhancing personalized fitness and health tracking.[9]

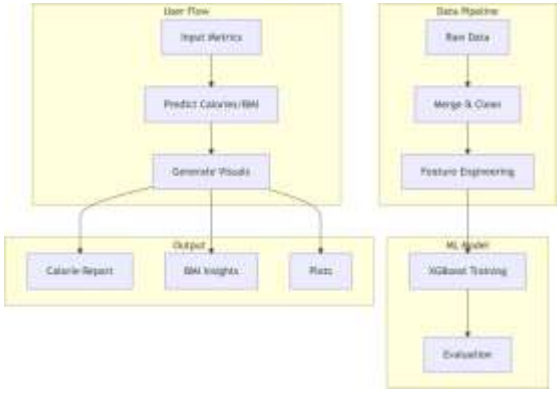
ACCURATE CALORIE BURN PREDICTION WITH MACHINE LEARNING: XGBOOST IN FOCUS

Gurrappagaru Sanjana Reddy, Singareddy Velankini Suhas

This study focuses on accurately predicting calorie expenditure using machine learning, with a particular emphasis on the XGBoost regression algorithm. Utilizing over 15,000 data points from two diverse datasets containing physical activity and biometric data, the research demonstrates the superior performance of XGBoost compared to other regression models. The predictions are based on established metabolic equivalent of task (MET) charts and formulas, highlighting the potential of machine learning in

supporting personalized health and fitness tracking.[10]

III. METHDODLOGY



A. Data Acquisition and User Input

The initial phase involves collecting demographic and physiological data from users. Key attributes include age, gender, height, weight, exercise duration, heart rate, and body temperature. Users input this information via a secure web or mobile interface, ensuring data privacy and integrity. This phase establishes the foundation for personalized predictions by capturing all relevant variables affecting calorie expenditure.

B. Data Preprocessing and Feature Engineering

Once raw data is collected, it undergoes a robust preprocessing pipeline to ensure consistency and quality. This includes:

- Handling missing or inconsistent values through imputation or validation checks.
- Encoding categorical variables (such as gender) for machine learning compatibility.
- Normalizing or standardizing numerical features to ensure uniform input scale.
- Calculating derived features such as Body Mass Index (BMI) using height and weight, providing additional context for calorie prediction.

This pre-processing ensures the dataset is clean, stable, and suitable for high- quality model training and inference.

C. Exploratory Data Analysis and Feature Selection

Before model training, exploratory data analysis (EDA) is conducted to understand the distribution of features and their relationship with the target variable (calories burned). Statistical techniques and visualizations are used to identify outliers, correlations, and feature importance. Based on these insights, relevant features are selected or engineered to maximize model performance.

D. Model Training and Calibration

The core of the system is the XGBoost Regressor, a robust gradient boosting algorithm known for its efficiency and predictive power in structured data scenarios. The model is trained on a curated dataset split into training and validation sets. Hyperparameter tuning is performed using cross-validation to optimize performance. The training process involves:

- Feeding preprocessed and engineered features into the model.
- Iteratively adjusting model parameters to minimize prediction error (e.g., Mean Absolute Error, Root Mean Squared Error).
- Comparing XGBoost's performance with baseline models (e.g., Linear Regression, Random Forest) to validate its superiority for this task.

E. Calorie Prediction and Fitness Insights Generation

Upon successful training, the model is deployed to generate real-time calorie burn predictions based on new user inputs. In addition to calorie estimation, the system provides:

BMI classification (underweight, normal, overweight, obese).

Exercise duration estimation required to burn a user-specified number of calories.

Weight loss projections based on caloric deficit principles (e.g., 3500 calories ≈ 1 pound of fat). These insights are presented through an interactive dashboard, enabling users to track progress and adjust fitness plans accordingly.

F. System Interface and User Roles

The platform features two primary modules:

User Module: Allows users to register, log in, input personal and exercise data, and view personalized calorie predictions and fitness recommendations. Users also have access to curated health tips and exercise information.

Admin Module: Enables administrators to manage user accounts, curate and update health tips, and maintain the exercise database, ensuring the platform remains current and informative.

G. Real-Time Prediction and Output Display

When a user submits new data, the system processes the input through the preprocessing and feature engineering pipeline, applies the trained XGBoost model, and instantly displays calorie burn estimates and tailored fitness insights on the user interface. This real-time feedback loop enhances user engagement and supports informed decision-making.

H. Evaluation and Testing

The system is rigorously evaluated using a diverse set of test cases representing various user profiles and exercise scenarios. Performance metrics such as MAE, RMSE, and user satisfaction scores are tracked to ensure robustness and reliability. Continuous monitoring and periodic retraining are implemented to maintain model accuracy as new data becomes available.

IV. TECHNOLOGIES USED

A. Python

Python is a high-level, general- purpose programming language known for its clean syntax and ease of readability. It is widely used across various domains such as web development, automation, data science, artificial intelligence, and software development. One of Python's greatest strengths is its extensive library support, which allows developers to work faster and more efficiently. The language is platform- independent and has a large and active community, making it easier for developers to find support, resources, and solutions to problems. Python's simple structure and dynamic typing make it an ideal choice for both beginners and professionals working on complex projects.

Python libraries are

B. Pandas

Pandas is a fast, powerful, and flexible open-source data analysis and manipulation library built on top of the Python programming language. It provides

intuitive data structures—most notably the DataFrame and Series—that make working with tabular, time series, and other labeled data both easy and efficient. Widely used in data science, machine learning, and real- world data analysis, pandas offers robust tools for loading, cleaning, transforming, and analyzing data, supporting operations such as merging, filtering, grouping, and handling missing values. Its versatility and performance have made it an essential tool for analysts and developers, enabling them to process messy or raw datasets into clean, structured formats ready for analysis and modelling.

C. XGBoost

XGBoost (eXtreme Gradient Boosting) is a highly efficient, open-source machine learning library that implements optimized gradient boosting algorithms using decision trees for both classification and regression tasks. Renowned for its speed, scalability, and predictive accuracy, XGBoost builds an ensemble of weak learner trees by sequentially correcting the errors of previous models, resulting in a robust and high-performing predictor. It offers advanced features such as regularization (to prevent overfitting), parallel and distributed computing, and the ability to handle missing values, making it a popular choice for data scientists in both industry applications and machine learning competitions.

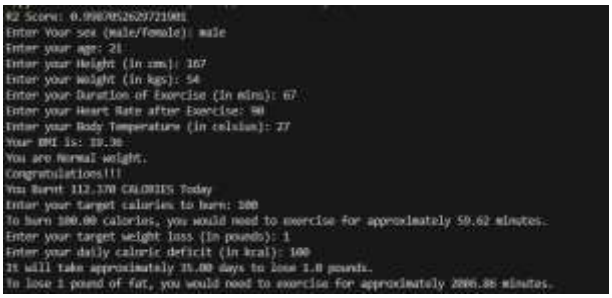
D. NumPy (Numerical Python)

NumPy is a core library in Python for numerical computing, particularly useful for working with large arrays and matrices. It provides a powerful N-dimensional array object along with a collection of mathematical functions to perform operations such as linear algebra, statistics, and Fourier transforms. NumPy is highly optimized for performance and is considered the foundation for many scientific computing libraries in Python, including pandas, SciPy, and scikit-learn. It is often used in data preprocessing tasks and integrates seamlessly with OpenCV and TensorFlow, making it essential for image processing, machine learning, and data science projects.

E. Sklearn

Scikit-learn, commonly known as sklearn, is a widely used open-source Python library that provides a comprehensive suite of tools for machine learning and data analysis. Built on top of core scientific libraries like NumPy and SciPy, scikit-learn features a consistent and user-friendly API for implementing a variety of machine learning algorithms, including classification, regression, clustering, dimensionality reduction, and model selection. It supports popular algorithms such as support vector machines, random forests, k-means, and gradient boosting, and is designed to integrate seamlessly with other Python libraries for data manipulation and visualization. Scikit-learn's robust documentation and simple fit/predict workflow make it accessible to both beginners and experienced practitioners, enabling efficient development, evaluation, and deployment of machine learning models across a range of real-world applications.

V Result



The system predicts calories burned with high accuracy (R^2 score) and provides actionable insights like BMI classification, exercise duration targets, and weight loss timelines. Visualizations clearly show how duration and intensity impact calorie expenditure, helping users optimize workouts.

VI. CONCLUSION

In conclusion, this paper, we presented a comprehensive machine learning-based system for calorie burn prediction and personalized fitness insights, leveraging user demographics and exercise-related data to deliver accurate, actionable recommendations. By integrating advanced preprocessing techniques, feature engineering, and the XGBoost Regressor model, our approach effectively addresses the challenges of individual variability and data complexity in calorie expenditure estimation. The platform's modular architecture, featuring distinct Admin and User roles, ensures both robust system management and an engaging user experience. Additional functionalities such as BMI classification, exercise duration estimation, and weight loss projections further enhance the practical value of the system for users pursuing diverse health and fitness goals.

Experimental results demonstrate that our model achieves high predictive accuracy, validating the effectiveness of machine learning for this application. The system's real-time feedback and intuitive interface empower users to make informed decisions, optimize their workout routines, and track progress towards their goals. As a holistic solution, this platform not only bridges the gap between generic fitness advice and personalized health management but also lays the groundwork for future enhancements, such as integration with wearable devices and expansion to broader health analytics. Ultimately, our work contributes to the advancement of intelligent, user-centric fitness technologies that support healthier lifestyles and improved well-being.

VII. REFERENCES

1. A Study on Calories Burnt Prediction Using Machine Learning, Authors: Punita Panwar, Kanika Bhutani, Rimjhim Sharma, Rohit Saini, Publisher: ITM Web of Conferences (EDP Sciences), DOI: https://doi.org/10.1051/itmconf/202354010 10

2. Calorie Burn Prediction: A Machine Learning Approach using Boost XG Regression, Authors: Niharikareddy Meenigea, Publisher: Journal of Emerging Technologies and Innovative Research (JETIR), eISSN: 2349-5162

3. A Study on Calories Burnt Prediction Using Machine Learning Punita Panwar, Kanika Bhutani, Rimjhim sharma and Rohit Saini, Publisher: ITM Web of Conferences 54, 01010 (2023) https://doi.org/10.1051/itmconf/202354010 10 I3CS-2023

4. MACHINE LEARNING

APPROACHES FOR CALORIE BURN ANALYSIS AND PREDICTION Neha Valmiki1 , Ambili P S, Publisher: EPRA International Journal of Multidisciplinary Research (IJMR), ISSN (Online): 2455- 3662, Article DOI: https://doi.org/10.36713/epra15281 DOI No: 10.36713/epra15281

5. Calorie Burn Predection using Machine Learning Suvarna Shreyas Ratnakar, Vidya S, Publisher: International Advanced Research Journal in Science, Engineering and Technology ISO 3297:2007, ISSN (O) 2393-8021, ISSN (P) 2394-15

6. Calorie Burnt Prediction using XGBoost & Exercise Duration, Dasam Mohan Murali Babu, Publisher: / International Journal of Engineering & Science Research, ISSN 2277-2685 IJESR/April-June. 2025/ Vol-

15/Issue-2s/375-382

7. Comparing machine learning algorithms for predicting calories burned, Punlisher: Gudiwada vijayalakshmi,Torati Sridurga, Publisher: 2023 JETIR March 2023,
Volume 10, Issue 3 www.jetir.org (ISSN-2349-5162)

8. A Study on Calories Burnt Prediction Using Machine Learning Punita Panwar, Kanika Bhutani, Rimjhim sharma and Rohit Saini, Publisher: ITM Web of Conferences 54, 01010(2023) https://doi.org/10.1051/itmconf/20235401 010 I3CS-2023

9. Calories Burnt Prediction: A Machine Learning Approach Md Nahid Hosain Likhon, Farhan Bhuiyan, Md. Sihab Bhuiyan, Monjurul Aziz Fahim , and Amzad Hossain , Publisher: TechRxiv (preprint server) , The paper was made publicly available on August 26, 2024, and can be accessed using the DOI: 10.36227/techrxiv.172469977.72861155/v
1.

10. ACCURATE CALORIE BURN PREDICTION WITH MACHINE LEARNING: XGBOOST IN FOCUS
Gurrappagaru Sanjana Reddy, Singareddy Velankini Suhas, Publisher: International Research Journal of Modernization in Engineering Technology and Science, e- ISSN: 2582-5208,) www.irjmets.com