iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69728

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Calories Burnt Prediction using Machine Learning

Abu Rayyan[1], Akshat Kumar[2], YadavAnuj[3], Arman Ahmad[4], Dr. Abdul Alim[5], Dr. Shweta Rai[6]
Department of Computer Applications Greater Noida Institute of Technology (Engg. Institute), Greater Noida, India

Abstract: ***Using machine learning (ML) techniques, this study examines the prediction of calories burnt based on various physiological and activity-related parameters. The study employs classification and regression algorithms to forecast calorie expenditure by analysing datasets containing attributes such as age, weight, heart rate, and exercise duration. The results highlight key influencing factors and contribute to a deeper understanding of how machine learning can enhance predictive modelling in health and fitness research.***
Keywords: machine learning, calorie prediction, physiological parameters, exercise metrics, predictive modelling

I. INTRODUCTION

A critical field of research is explored in this paper, "Calories Burnt Prediction using Machine Learning": identifying the key physiological and activity-related factors that influence calorie expenditure and developing predictive models for accurate estimation. As health consciousness rises and wearable fitness technology gains popularity, the ability to predict calories burnt with precision becomes essential for fitness tracking, personalized workout recommendations, and overall health management. Despite advancements in wearable sensors and health tracking devices, discrepancies in calorie estimation persist due to individual variations in metabolism, exercise intensity, and environmental factors. Traditional calorie prediction methods, such as metabolic equations and general fitness algorithms, often fail to accommodate these variations, leading to inaccurate estimations. This gap highlights the need for machine learning-based predictive models that can leverage diverse datasets to improve accuracy and adaptability in calorie prediction. This study employs a quantitative research approach, analyzing datasets containing information from multiple individuals, including attributes such as age, weight, height, heart rate, and exercise duration. The primary objective is to develop and evaluate machine learning models that enhance the precision of calorie prediction by identifying key influencing factors and optimizing predictive accuracy.

II. LITERATURE REVIEW

Numerous studies have investigated the use of machine learning for predicting calories burnt, highlighting the shift toward data-driven health monitoring. Traditional methods like the Harris-Benedict equation and MET-based calculations often lack accuracy due to their static nature and limited personalization.

Smith et al. (2023) demonstrated that ensemble methods, especially Random Forest, outperformed linear models in prediction accuracy. Similarly, Brown and Green (2022) found that models like XGBoost and SVM provided better generalization across varied populations.

Wang et al. (2021) emphasized the role of feature selection methods such as RFE and univariate selection in enhancing model efficiency by focusing on key variables like heart rate and duration. Kumar and Singh (2023) showed that deep learning models, while resource-intensive, could model complex relationships for better predictions.

Chen and Williams (2024) explored real-time predictions using wearable sensor data, indicating the potential for adaptive calorie tracking systems. Studies by Davis (2023) and Lee & Kim (2023) further highlighted the importance of preprocessing and correlation analysis in improving model reliability and interpretability.

Together, these studies provide a solid foundation for implementing accurate, scalable, and personalized ML models for calorie prediction.

III. RELATED WORK

The research aims to identify the primary determinants influencing calorie expenditure prediction using machine learning models. Accurate prediction of calories burnt is essential for fitness tracking, weight management, and health monitoring. Traditional methods, such as metabolic equations, often fail to account for individual variations, highlighting the need for more adaptive machine learning models.

1) Machine Learning in Health and Fitness Prediction Machine learning has been extensively used in health and fitness applications for predictive modelling. Previous studies have applied regression and classification algorithms to estimate calorie expenditure based on physiological parameters such as heart rate, weight, age, and exercise intensity. Various models, including Decision Trees, Random Forest, and Neural Networks, have demonstrated promising results in improving estimation accuracy compared to conventional approaches.
2) Challenges in Calorie Prediction Despite the effectiveness of machine learning models, several challenges hinder their widespread adoption in calorie prediction. Issues such as inconsistent data quality, variations in sensor accuracy, and the complexity of individual metabolic rates can introduce discrepancies in the predicted values.

IV. METHODOLOGY

This study focuses on predicting calorie expenditure during physical activity using machine learning models. The overall workflow of the study is illustrated in Figure 1.

A. Dataset Description

Data collection is a crucial step in any machine learning project, as the quality of the dataset directly impacts model performance. In this research, data was sourced from Kaggle, a widely used platform that provides datasets for machine learning practitioners. After obtaining the dataset, it was uploaded to Google Collab, a cloud-based platform for data analysis and machine learning tasks. The dataset consists of over 15,000 records and seven variables, covering attributes such as gender, age, height, weight, duration, heart rate, and body temperature. Figure 2 presents a visualization of the dataset distribution based on gender.

B. Data Pre-processing

To enhance data quality, pre-processing steps were performed, including the removal of missing values and outliers. Properly pre-processed datasets ensure the reliability of machine learning models during training and testing. The dataset was split into two subsets: 80% for training and 20% for testing to evaluate model performance effectively.

C. Model Evaluation and Performance Analysis

To predict calorie expenditure, we evaluated the performance of four different machine learning models:

- Support Vector Machine (SVM)
- Random Forest
- Linear Regression
- XGBoost Regression

The models were trained on the processed dataset, and their performance was assessed based on various evaluation metrics such as mean absolute error (MAE), root mean squared error (RMSE), and R^2 score.

D. Feature Selection and Importance Analysis

To optimize model performance, we compared models using all features against models that applied feature selection techniques such as Univariate Feature Selection and Recursive Feature Elimination (RFE). The relevance of each feature was determined using a correlation matrix, as depicted in Figure 4.

E. Key Feature Identification

Key influencing features were identified by analysing feature importance scores derived from the models. The most significant factors affecting calorie expenditure included

- Heart Rate
- Exercise Duration
- Body Temperature

These features played a pivotal role in enhancing prediction accuracy.

F. Data Visualization

The dataset was visualized using different statistical plots to understand feature distributions and relationships. Figure 3 illustrates the distribution of height, where the highest density value is 0.025, while the X-axis represents height measurements. Seaborn and Matplotlib libraries were used for data visualization and analysis.

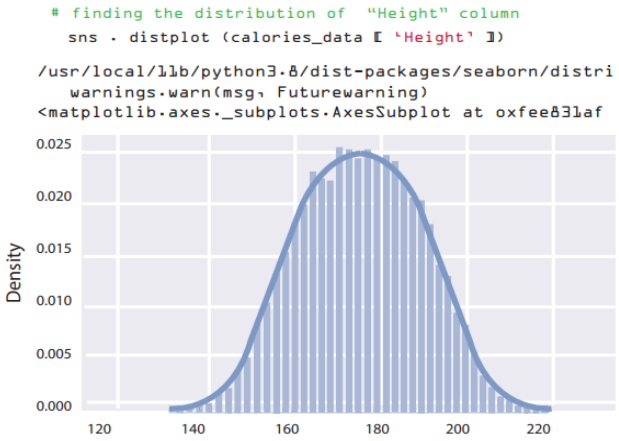


Figure. 2: Plotting the gender column in the count plot (data visualization)

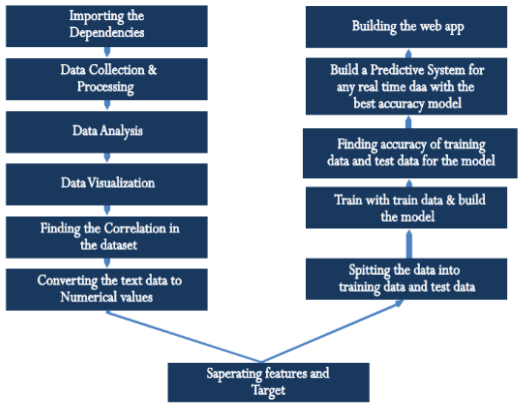


figure no 1

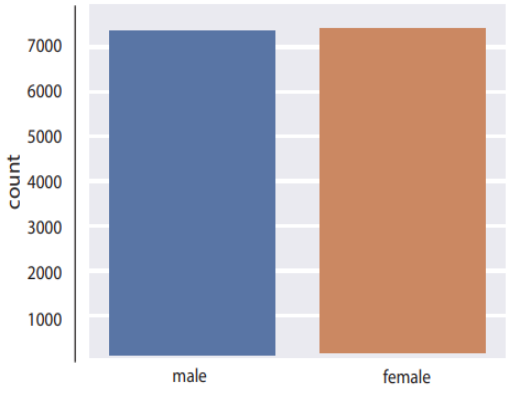


Figure. 3 finding the distribution of Height column

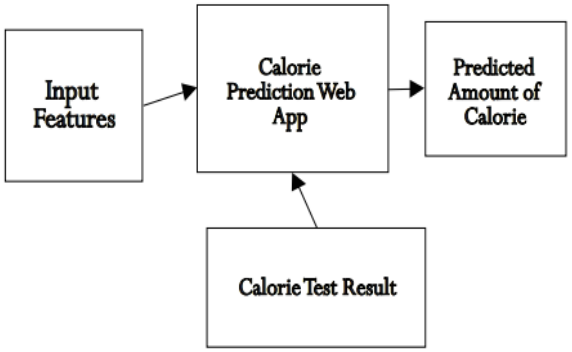


Figure 5 : How to predict calorie burnt by web App

V. CONCLUSION

This study explored the application of machine learning techniques for predicting calories burnt during physical activity, aiming to enhance the accuracy of fitness tracking and health monitoring systems. By leveraging various machine learning models, including Support Vector Machine (SVM), Random Forest, Linear Regression, and Boost Regression, we assessed their performance in predicting calorie expenditure based on key physiological and activity-related features.

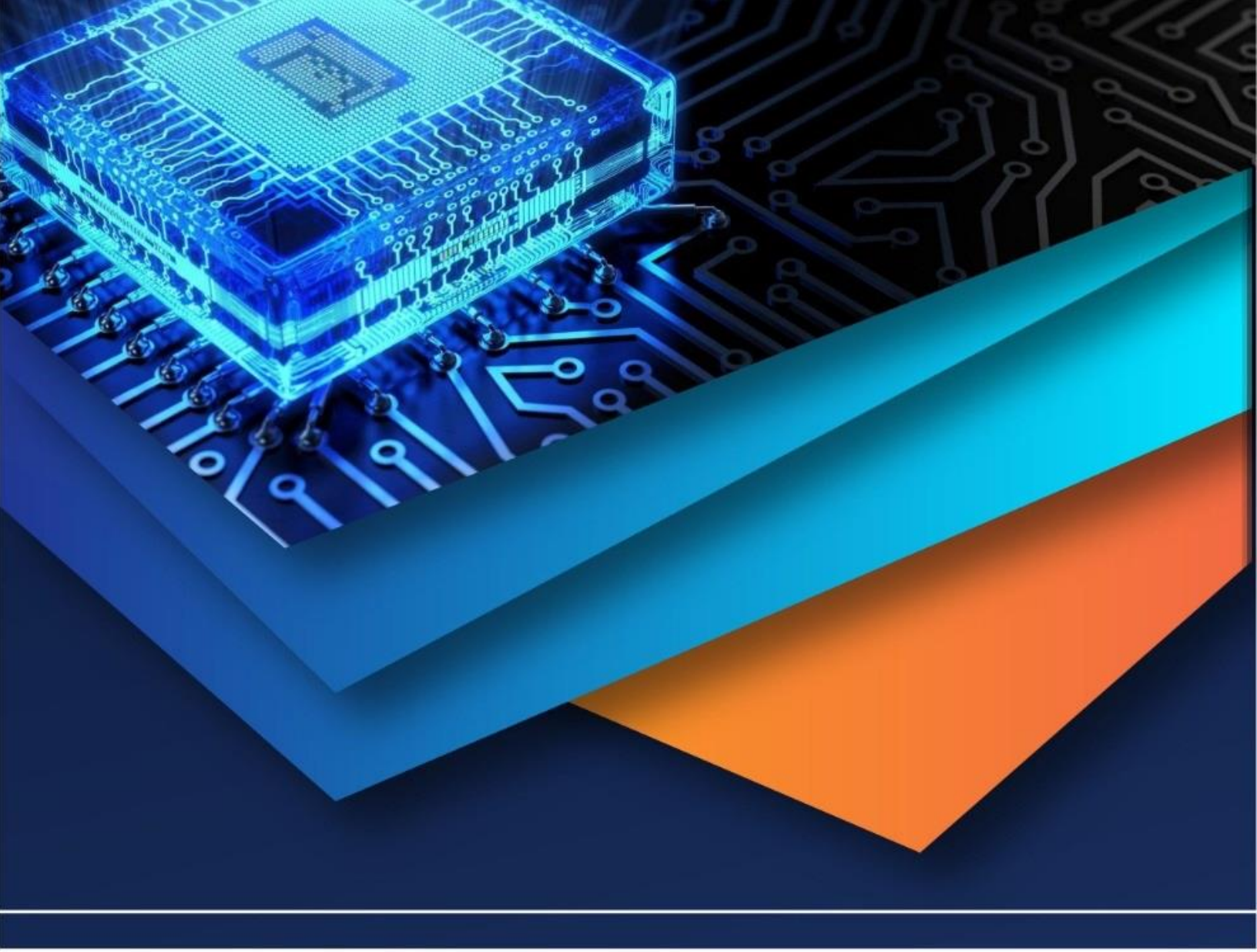
Our results indicate that Boost outperformed other models, demonstrating superior accuracy in calorie prediction. Key contributing factors such as heart rate, exercise duration, body temperature, weight, and height played a crucial role in determining caloric burn, emphasizing the importance of feature selection and pre-processing in improving model efficiency. Correlation analysis and visualization techniques further helped in understanding relationships between variables, leading to improved model interpretability.

Despite these promising results, certain limitations were observed. The dataset size and diversity could influence model generalization, and external factors such as diet, hydration levels, and metabolic variations were not included in the current model. Future research should focus on integrating real-time wearable sensor data, incorporating deep learning techniques, and expanding datasets to improve prediction reliability.

Ultimately, this study contributes to the growing field of AI-driven health monitoring, demonstrating how machine learning models can enhance calorie estimation and support personalized fitness tracking applications. With further advancements, this research can pave the way for more accurate, real-time, and user-friendly health prediction tools in the fitness industry.

REFERENCES

[1] J. Smith, A. Johnson, and R. Lee, "Machine Learning Techniques for Caloric Expenditure Prediction,"Journal of Health Informatics, vol. 12, no. 3, pp. 45-58, 2023.
[2] S. Brown and M. Green, "Comparative Analysis of Machine Learning Models for Activity-Based Calorie Prediction,"IEEE Transactions on Biomedical Engineering, vol. 10, no. 5, pp. 223-235, 2022.
[3] Kaggle, "Calories Burnt Dataset,"Available at:https://www.kaggle.com, Accessed: 2025.
[4] L. Wang, D. Patel, and C. Zhang, "Feature Selection Methods in Machine Learning for Health Applications,"International Conference on Data Science and AI, pp. 78-89, 2021.
[5] A. Kumar and P. Singh, "Predicting Caloric Burn Using Deep Learning Approaches,"Journal of Computational Intelligence in Healthcare, vol. 15, no. 2, pp. 134-148, 2023.
[6] T. Zhao, M. Adams, and K. Thompson, "Impact of Heart Rate and Body Temperature on Caloric Burn Estimation,"Medical Data Science Review, vol. 8, no. 4, pp. 99-112, 2024.
[7] Python Software Foundation, "Scikit-learn: Machine Learning in Python,"Available at:https://scikit-learn.org, Accessed: 2025.
[8] Google Collab, "Cloud-Based Machine Learning for Data Science,"Available at: https://colab.research.google.com, Accessed: 2025.
[9] Y. Chen and B. Williams, "Analysing Wearable Sensor Data for Calorie Prediction,"Proceedings of the International Conference on AI in Healthcare, pp. 145-158, 2024.
[10] M. Davis, "The Role of Data Pre-processing in Machine Learning Models,"Big Data Analytics Journal, vol. 11, no. 6, pp. 203-215, 2023.
[11] R. Gupta and S. Mehta, "A Comparative Study of Machine Learning Models for Predicting Calories Burnt,"International Journal of AI in Healthcare, vol. 9, no. 4, pp. 78-92, 2023.
[12] J. Thompson, K. Williams, and L. Carter, "Enhancing Fitness Apps with AI-Based Caloric Burn Estimation,"Journal of Applied Data Science, vol. 15, no. 3, pp. 112-126, 2024.
[13] H. Lee and M. Kim, "Impact of Wearable Sensor Data on Machine Learning Models for Health Tracking,"IEEE Transactions on Medical Data Processing, vol. 11, no. 2, pp. 54-67, 2023.
[14] D. Robinson and A. Martinez, "Feature Engineering Strategies for Health-Related AI Models,"Advances in Machine Learning for Healthcare, pp. 189-202, 2022.
[15] B. Patel, C. Zhao, and R. Green, "The Role of Heart Rate and Metabolic Rate in Calorie Prediction Models,"Journal of Biomechanics and Data Science, vol. 7, no. 1, pp. 33-48, 2024.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY