Received: May 18, 2024. Revised: May 24, 2024.

Calories burnt Prediction Using Machine Learning Algorithms

-Mosaab Emad 202000092
-Emad Sakr 202001824
-Mohamed Yasser 202000430
-Omar Ayman 202000537
-Marwa Ali 202000853

Nile University

Abstract: Calories is essential for human being to consume throw the day. So, we predict the number of calories need for the individual to consume to run his/her system which is called basal metabolic rate (BMR) using machine learning algorithms by knowing the age, gender, height, and weight.

1. Agenda

2. Introduction

Understanding and predicting calories burned, particularly through Basal Metabolic Rate (BMR), is crucial for numerous health and fitness applications. BMR represents the number of calories required to maintain basic physiological functions at rest, such as breathing, circulation, and cellular production. Accurate prediction of BMR is essential for tailoring individualized dietary and exercise plans, thereby aiding in weight management, athletic performance optimization, and the prevention of obesity-related conditions. This paper explores advanced methods for predicting BMR, integrating data from various demographic and physiological parameters. By leveraging machine learning algorithms and statistical models, we aim to enhance the accuracy of BMR predictions beyond traditional approaches, which often rely solely on age, gender, weight, and height. This improved precision is vital for creating effective diet plans that align caloric intake with personal metabolic needs, ultimately leading to more successful health interventions and better nutritional management.

3. Problem statement

The accurate prediction of daily burnt calories is a crucial component in personalized health and fitness management. This study aims to develop a robust machine-learning model that leverages individual-specific data to forecast daily calorie expenditure. Key input variables are age, gender, height, and weight each of which plays a significant role in determining metabolic rate and energy expenditure. Understanding the precise amount of calories burnt each day is essential for various applications, including weight management, nutritional planning, and the optimization of physical training. In addition to predicting daily calorie expenditure, we also calculate the body mass index (BMI) to classify wither the person is underweight, fit, or overweight.

4. Data size and description

We collected our data from many resources which are 2 gyms (club13 gym and H2O gym), our friends, and collected some data from the inbodies that was published from user to ask for a review in social media platform.

```
df.shape
(200, 7)
Age  Gender  Height  Weight  BMR  BMI  Category
```

Total Instances:200

Number of features: 7

Target Variable: 3

Feature Description:

Age: in years

Gender: male or female

Height: in centimetre

Weight: in kilograms

Target Variable:

BMR: basal metabolic rate in Kcal

BMI: body mass index

Category: skinny, fit, and fat

Outcome: amount of calories burnt (BMR)

5. Literature Review

The other papers conducted studies that aims to predict the calorie burn during physical activity using machine-learning models. By knowing the follows the duration of the workout, the heart rate, the body temperature, age, height, and weight. Our goal is to predict the maintenance calories of the day (BMR) for everyone. We also calculated the BMI of the individual to classify wither they are underweight, fit, or overweight.

6. Model Plan Architecture

We first imported necessary libraries and loaded the dataset, then visualized gender distribution.
The dataset underwent pre-processing before training the machine learning models to optimize performance and stability. The data is split by gender (male and female) and we added a new feature, which is body mass index (BMI) calculated by using weight and height. We built four models, which are linear regression, multilinear regression, decision tree, and random forest.
First the data set was split in to 2 groups males and females then the data was split in 2 group train group and test group with a ratio of 80% to 20%.

Linear Regression Model:

The equation of the linear regression is as follows $y=\beta_0+\beta_1 X$ we use only one feature in order to get the target. The feature is area WxH and the target is the BMR.

Multilinear Regression Model:

The equation of the linear regression is as follows $y=\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_n X_n$ we use multiple features in order to get the target. The features are age, height, and weight and the target is the BMR.

Decision Tree Model:

It works by splitting the data into subsets based on the value of input features (age, height, weight), creating a tree-like structure of decisions. Each internal node represents a decision based on a feature, each branch represents the outcome of that decision, and each leaf node represents a final output (BMR).

Random Forest Model:

It works by constructing multiple decision trees during training and outputting the mean prediction of the individual trees.

7. Evaluation metrics and results

In each model, we split data by gender and prepare training and testing sets and test size is 20% of the data. We evaluate the model using mean square error (MSE) and R2 score.

Linear Regression Model

Male MSE	2893.241520124566
Male R2 score	0.915733240807177
Female MSE	3703.275757545612
Female R2 score	0.786570771537137

Multilinear Regression Model

Male MSE	1939.3328576376546
Male R2 score	0.9435161932589837
Female MSE	3191.2348804953652
Female R2 score	0.8160809934285554

Decision Tree

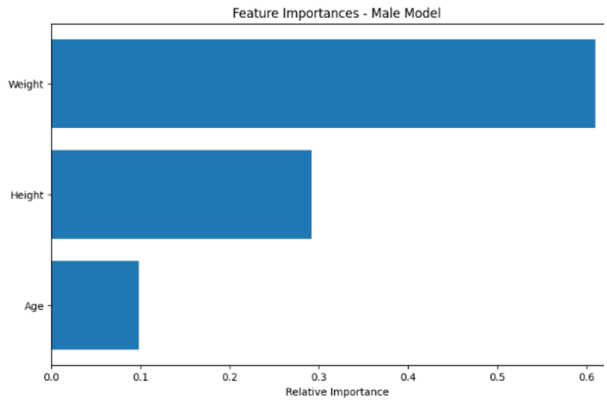
Male MSE	2299.7894736842104
Male R2 score	0.9330177572844108
Female MSE	3052.2542321124456
Female R2 score	0.8240910466231328

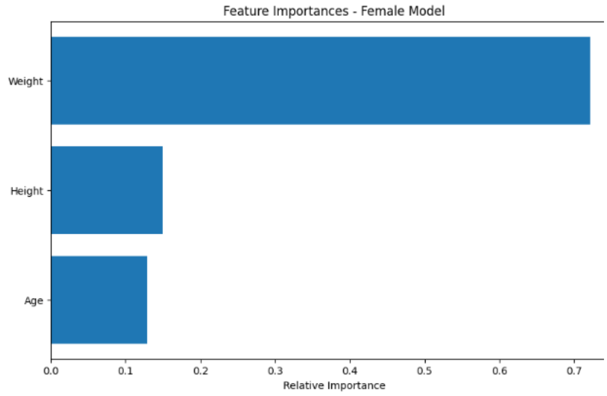
Random Forest

Male MSE	1785.8740631578962
Male R2 score	0.9479857389875405
Female MSE	3142.0245166666646
Female R2 score	0.8189171122249848

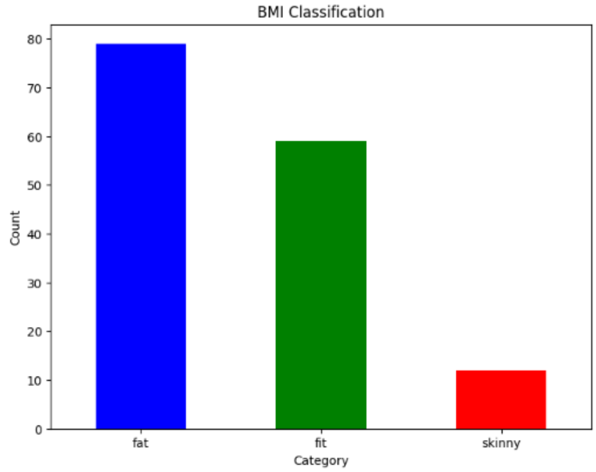
The most important feature for male and female is found to be weight.

Received: May 18, 2024. Revised: May 24, 2024.





BMI classification based on our dataset.



8. Model selection

Based on the evaluation metrics and results, the Random Forest Regression model performed the best for male model as it achieved the lowest mean square error and highest R2 score. For female model, the decision tree performed the best as it achieved the lowest mean square error and highest R2 score than the other three models.

9. Conclusion

In conclusion, our target is predicting the BMR. We relied on age, gender, and height. We developed and evaluated four model linear regression, multilinear regression, decision tree, and random forest using a dataset collected from two gyms, friends, and social media platforms. The dataset comprised 200 instances with features including age, gender, height, and weight. Additionally, we calculated the Body Mass Index (BMI) to classify individuals as underweight, fit, or overweight. Our findings reveal that the Random Forest Regression model yielded the best performance for the male subset, achieving the lowest mean square error (MSE) and highest R2 score. For the female subset, the decision tree model outperformed the others. These results underscore the importance of tailored models based on gender to enhance predictive accuracy.

10. Future Work

Collect more data from more resources

Expanding the dataset to further enhance the accuracy prediction of Basal Metabolic Rate (BMR).

Additional predictive models

Trying a different models can make us find a better accuracy and results.

Adding new features

We can add new features that help us make a more accurate such as prediction body fat percentage, muscle mass, activity level, and dietary habits.

References

Data was collected from club 13 and h2o gym.

https://www.geeksforgeeks.org/calories-burnt-prediction-using-machine-learning/