

A Study on Calories Burnt Prediction Using Machine Learning

Punita Panwar¹, Kanika Bhutani², Rimjhim sharma³ and Rohit Saini⁴

^{1,2,3,4}Jaipur Engineering College & Research Centre, Computer Science Department, Jaipur, Rajasthan, India

Abstract. In this growing technological era, People are less aware of their health and mental stability. Due to lack of time, they intake more junk food than healthy options, which leads to an increase in the total calorie rate in their body. which is a major cause of obesity a calorie is the rate of energy stored and energy expenditure. People nowadays want quick solutions to every problem they want to exercise less and get more results, so to check the level of improvement and the burnt calories level after exercise in the human body we came up with this machine Learning System which takes some attributes as input and gives approximate calories burnt value which will motivate people to do more exercise and will show their daily growth The project is nourished with more than 15,000 data and its MAE (Mean Absolute error) is 1.48 which will enhance over time for better Results.

Keywords- ML(Machine learning), Kaggle, Collab, XGBOOST, Decision tree, linear Regression, SVR, AdaBoost regressor.

1 Introduction

The amount of calories burnt depends on internal and external factors, it is subjective and different for everyone depending on their height, weight, and fitness level [1]. Generally, people relate calories to weight or food reduction however it is a quantity of heat energy. From a human perspective, the number of calories is the amount of energy required to carry out a task. Different items have different calorie values related to them.

As a human body performs some extensive activity or workout the body temperature and heart rate start rising which leads to the production of heat energy in the body. Which ultimately causes calories to burn. To show the same we take some input parameters such as age, gender, height, and weight and apply different regression algorithms such as linear regression, XG Boost regression, AdaBoost regression, SVR, Decision tree regression, and Random forest regression over the data to get the best and optimal results.

Factors affecting calories burnt during exercise are as shown in fig-1-

1. Age (in years) –As age increases the need for calories decreases in the body. This is often due to less physical activity, changes in metabolism, or age-related loss of bones and muscle mass.
2. Height- Taller people have greater lean mass, which is related to metabolic rate .which implies that people need more calories to function. Hence the calorie rate in taller people will be more than in shorter ones.
3. Weight- If you intake more calories than the calories you burn your weight will start rising. it means the more the weight of a person the more they need to burn calories to balance their intake calories.
4. Gender- Human body's calorie needs vary according to gender, women need fewer calories than a man Because men have greater muscle mass.
5. Heart rate- While exercising heart rate increases and shows the exertion in doing an activity which determines the calories you burn. Also, a higher heart rate leads to body fat burn which is a composition of calories.
6. Body Temperature- As body temperature rises, more energy is used to cool it down which leads to calorie burn.
7. Duration- With the increase in the duration of exercise body temperature also increase, which leads to extra calorie burn hence the duration of exercise directly affects the calories burned in the body.

¹Punita Panwar: punitapanwar7@gmail.com

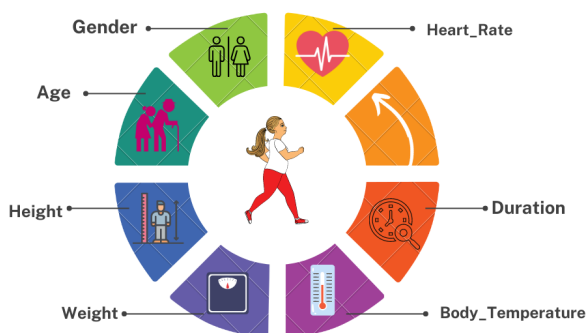


Fig-1 Attributes of calories burnt prediction

2 Literature Survey

Daniel Bubnis [3] The number of energy burned in everyday life is depending upon weight, height, age, and gender. Amount of calories people need to burn then they eat, cause a lack of calorie. But it is really important to understand what amount of calories they burn day-to-day life. Calories are the unit of heat or energy that is required to elevate 1 g (gram) of water to 1 c (Celcius). At the time of working amount calories are burnt day-to-day, so it is essential for a person trying to maintain their body.

Salvador Camacho [4] universal obesity has been increasing day by day across the whole world and until now not a single nation has been able to resolve it. The main cause of obesity is an energy imbalance between calories eaten and calories expended. The concept of calorie imbalance can not be sufficient to control and turn the obesity pandemic.

The World Health Organization (WHO) [5] There are many factors that affect the calories burned, but anyone can be modified their diet chart or activity level to get the desired results. There is a study in the literature that used ml and data mining to diagnose problems. When we compare from today's scenario some articles are published earlier with low accuracy of calories burned prediction problems.

Jadhav Kalpesh et al.[6] discussed about the prediction of human activity by considering mobile sensor data and they used LSTM and Neural Network for predicting human based activity.

Akshith Rajesh Tayadeet al. [7] used logistics regression algorithm for diet recommendation system to support mental fitness and physical fitness and accuracy of the proposed model was 85.96.

3 Methodology

This study gathers the right data set and trains the ml model to determine how many calories are burnt by an individual. Firstly, pre-processing of the dataset is carried out to make data free from null values and keep the important attributes in the dataset. After this data is plotted into different graphs to understand the relationship among attributes using different visualization Techniques. We used different regressive and linear machine learning algorithms to compare and find optimal solutions. This study shows the variance among results of different approaches using graphs.

Regression is the technique for investigating the relationship between variables and outcomes. It is used in predictive modeling in which algorithms are used to predict continuous outcomes.

Linear regression:- this model consists of a predictor variable and a dependent variable related linearly to each other. It is used to find the dependency between two variables. It calculates the amount of increase in temperature with the amount of exercise done.

¹Punita Panwar: punitapanwar7@gmail.com

XG boost regression:- It stands for extreme gradient boosting. This ensemble learning model produces strong predictions by combining the prediction of multiple weak models. It handles large datasets and provides efficient handling of missing values.

Ada boost regression:- also called adaptive boosting is used as an ensemble model.it is a statistical classification meta-algorithm.it is used to combine weak base learners to create an accurate model.

Decision tree regression:- It builds a regression model in a tree structure. To start constructing we start with a feature that will become the root node. the feature with the least impurity is considered as a node at any level sort the data in ascending order and calculates the average of adjoining values. Then the impurity at each level is calculated.

SVM(support vector machine):- it tries to find a hyperplane that separates two classes and then classifies a new point depending on whether it lies on the positive or negative side of the hyperplane depending on the class to predict.[8]

Random forest regression:- it is a supervised learning algorithm this model combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model. It is powerful and accurate.

The workflow of the project looks like as shown in fig[2]-

1. Data collection-dataset collection is the primary step. we used Kaggle as the data repository. Data is then uploaded to the collab platform. The data used here is both categorical and numerical.
2. Pre-processing of data- it is important that we process our data before passing it to the model for better results. null values and missing values are handled at this point because the information on our data directly affects how our model learns.
3. Analysis of data- firstly the two CSV files(“exercise.csv”, and “ calories.csv”) from Kaggle are uploaded to our used platform collab. Data visualization is carried out using various charts and graphs. the two types of correlation positive and negative are studied between various features. The data is then split into test and training data. the used regression models are loaded. test data is used to assess the prediction.
4. Machine learning model- all the chosen algorithms are applied at this stage to determine the r^2 value and absolute mean error value. Among the various algorithms, the best results are shown by XGBoost regression which gives the least absolute error value of 1.48 and efficient way to predict calories burnt.
5. Evaluation – the results of different algorithms are compared and the best among them is used to calculate the prediction of calories burnt during exercise along with various other factors such as age, gender height, weight, body temperature, and heart rate.

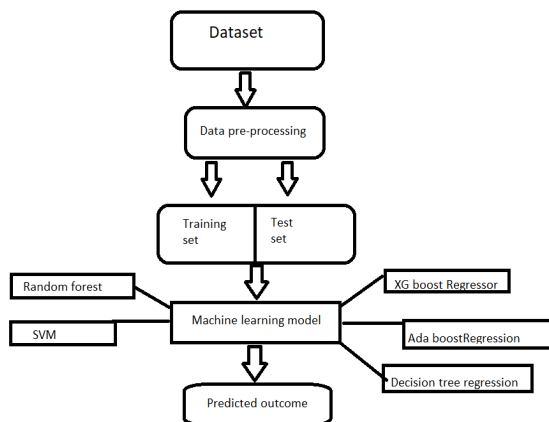


Fig 2- Workflow of machine learning model

¹Punita Panwar: punitapanwar7@gmail.com

4 Data source

We used the “Kaggle” Repository as our dataset store. We used two CSV files which hold 15000 instances for 7 attributes. The attributes range from Gender, Age, height, weight, Body temperature during exercise, Heart rate, and duration of workout in the “exercise.csv” file which is used as training data, “calories.csv” contains the corresponding values of calories burnt by individuals of the exercise data set.

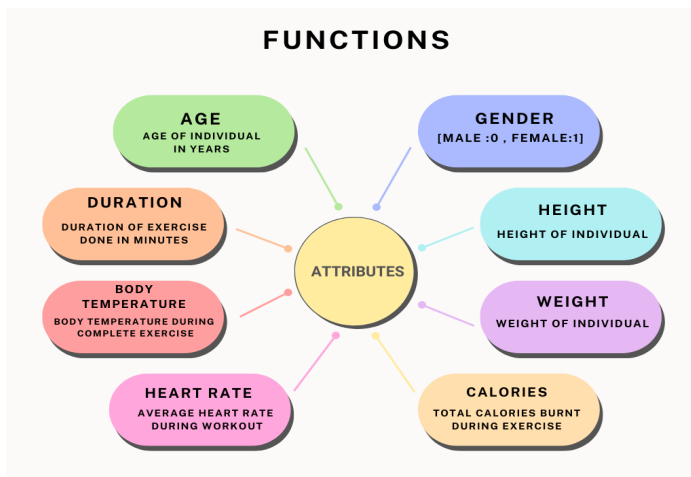


Fig3-Functions of different attributes of the dataset

5 RESULT

The analysis of this model is done to find the best algorithm for predicting the calories burnt during exercise from factors such as age, height, weight, body temperature, gender, heart rate, and duration of exercise. The algorithm which provides the least mean absolute error is considered as best, this study applies various machine learning models over the dataset to find the least value of Mae, according to these results XGBoost regression is best for solving this problem with a Mae value of 1.48. And the highest Mae value is of support vector regression which is 10.620 as shown in Table1.

Regressions	Mean Absolute Error (MAE)
Linear Regression	8.3851
Decision Tree Regression	3.343
Random Factor Regression	1.675
Ada boost Regression	8.670
XG boost Regression	1.4807
SVR	10.620

Table 1- Mean absolute error of different algorithms.

¹Punita Panwar: punitapanwar7@gmail.com

The difference between the results of various models is plotted using a visualization technique as shown in fig-4 for better understanding, the largest bar denotes the highest value of Mean absolute error and the lowest value of the Bar graph denotes the lowest value of Mae which is the required result of the research.

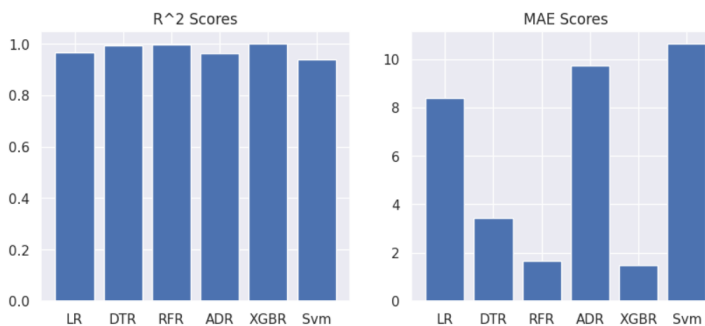


Fig4- Plotting graph of different MAE values and R² values

6 Conclusion

This research aimed to recognize the number of calories our body burns, which depends on several factors such as age, gender, weight, height, body temperature, duration, and heart rate. It is important to understand the number of calories we eat to stay fit and healthy. Calories burnt can be predicted from different regression algorithms such as Linear regression, XG boost regression, Ada boost regression, Decision tree regression, SVM, and Random forest regression. Out of these regression algorithms, Extreme Gradient Boosting (XG boost) regression gives the best accurate result. The MAE(Mean Absolute Error) value of the XG boost is 1.48 which is a good value. It means the errors are quite low. So, therefore, XG boost regression algorithm is the optimal algorithm for the calories burnt prediction so far.

References

1. Goukens, Caroline, and Anne Kathrin Klesse. "Internal and external forces that prevent (vs. Facilitate) healthy eating: Review and outlook within consumer Psychology." *Current Opinion in Psychology* (2022): 101328.
2. Khan, Abdul Wahid, et al. "Factors Affecting Fitness Motivation: An Exploratory Mixed Method Study." *IUP Journal of Marketing Management* 21.2 (2022).<https://www.medicalnewstoday.com/articles/319731>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5496172/>
4. Roberts, K. C., Shields, M., de Groh, M., Aziz, A., & Gilbert, J. A. (2012). Overweight and obesity in children and adolescents: results from the 2009 to 2011 Canadian Health Measures Survey. *Health rep*, 23(3), 37-41.
5. Kalpesh, Jadhav, et al. "Human Physical Activities Based Calorie Burn Calculator Using LSTM." *Intelligent Cyber Physical Systems and Internet of Things: ICICI 2022*. Cham: Springer International Publishing, 2023. 405-424.
6. Tayade, Akshit Rajesh, and Hadi Safari Katesari. "A Statistical Analysis to Develop Machine Learning Models: Prediction of User Diet Type."
7. Gour, Sanjay, et al. "A Machine Learning Approach for Heart Attack Prediction." *Intelligent Sustainable Systems: Selected Papers of WorldS4 2021, Volume 1*. Springer Singapore, 2022.

¹Punita Panwar: punitapanwar7@gmail.com

8. Panwar, Punita, et al. "A Prospective Approach on Covid-19 Forecasting Using LSTM." 2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP). IEEE, 2022.
9. Smola, Alex, and S. V. N. Vishwanathan. "Introduction to machine learning." Cambridge University, UK 32.34 (2008): 2008.'
10. Nipas, Marte, et al. "Burned Calories Prediction using Supervised Machine Learning: Regression Algorithm." 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T). IEEE, 2022.

¹Punita Panwar: punitapanwar7@gmail.com