

# Calorie Burn Prediction Analysis Using XGBoost Regressor and Linear Regression Algorithms

<sup>1</sup>Sona P Vinoy

Department of Computer Application  
Amal Jyothi College of Engineering  
Kottayam, India  
[sona123akd@gmail.com](mailto:sona123akd@gmail.com)

<sup>2</sup>Binumon Joseph

Asst.prof Department of Computer Application  
Amal Jyothi College of Engineering  
Kottayam, India  
[binumonjosephk@amaljyothi.ac.in](mailto:binumonjosephk@amaljyothi.ac.in)

**Abstract--The overarching idea of this research project is to make a comparative study of machine learning algorithms to predict the calories burn during the workout. In this paper we first build a machine learning systems that can predict the amount of calories burnt during exercise. In today's world many people are inquisitive about the workout that they do and the weight loss plan that they take and how much calorie do they burn once they workout. To solve this problem we can use ML algoirthms such as XGBoost regressor and Linear Regression.**

Keywords — Colab, XGBoost Regressor, Linear regression, machine learning, accuracy

## I. INTRODUCTION

The body temperature and the heartbeat will rise when we exercise or workout. The variables that we take here such as time scale for which the individual carrying out the workout training and what is the average beats per minute and then the temperature. Then we additionally take the height, weight, gender and age of the person to predict how tons energy the person may be burning.

A machine learning XGBoost regressor algorithm and linear regression algorithms are used to predict calories burned depends on the workout duration, body temperature, height, weight and age of the person.

## II. DESCRIPTION

The variety of energy burned each day is immediately connected to weight loss, weight gain, or weight maintenance. To shed pounds, a person ought to burn greater calories than they take in, developing a calorie deficit. but, to do that, they want to recognize what number of calories they burn each day. Most people think about calories as most effective having to do with food and weight reduction. Calorie, a unit of energy or heat variously defined. Calorie may be defined as the amount of energy that is vital to increase 1 gram(g) of water by means of 1 °C. This measurement can be carried out to lots of different

strength releasing mechanisms outdoor of the human body. In the case of human body, calories are measure of how much energy the body requires to function.

In order to be able to exercise how a whole lot of calories are burned every day is important to any man or woman trying to preserve, lose, or maintain weight. Understanding what elements contribute to calorie burning can help a person regulate their diet or workout program to deal with the aim. There are many factors that affect how much calories a person burns each day. some of the elements that effect day by day calorie burn aren't in a person's manage at the same time as others may be changed. These factors include :

In case of age, the older a person is, the fewer calories burned per day. Gender: men burn greater energy than women. Quantity of daily activity: Body composition: those with more muscular tissues, burn more calories than people who've much less muscle. body size: larger people burn greater calories than smaller human beings, even at relaxation. Thermogenesis: that is the amount of strength our body uses to break down meals.

### Key facts

- Now it is usually diagnosed that a excessive-calorie food regimen with inadequate bodily pastime consequences in overweight or obesity, which in return should result in illnesses together with noninsulin-established diabetes, hypertension, cardiovascular sicknesses, endometrial cancer, and gallstones.
- Raised BMI is a major fear element for noncommunicable sick consisting of cardiovascular diseases (specially coronary heart sickness and stroke), that have been the principle cause of dying in 2012-diabetes, musculoskeletal disorders (in particular osteoarthritis – a as a substitute disabling degenerative joint problems).

DOI: 10.5281/zenodo.6365018

ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

### Machine Learning

Machine learning is a dimension of artificial intelligence that's in certain mark out as the software applications turn out to be more correct at predicting outcomes without being explicitly programmed to accomplish that. To predict the new output values historical data is used by these algorithms. Machine learning let the user to feed a computer algorithm, massive amount of information and then the computer will examine and make data-driven suggestions and decisions focusing on only the input data.

### III. PURPOSE OF THE STUDY

This document is to predict the calorie burned during the workout of different people and compare the two algorithms in machine learning by looking through the data sets. The dataset used in this study has 7 features, one target variable, and 15000 instances. We are using this data sets to train a dataset and find out the accurate algorithms and its mean absolute error and find the best model.

### IV. METHODOLOGY

This paper is all about the collection of appropriate set to teach our machine learning models in order that it will find out what is the amount of calories that the individual goes to burn. Before feeding procedure the statistics via records pre-processing need to be done. After that data analysis is carry out where we use some visualization techniques to arrange the data in plots and graphs. Afterwards divide the data set into training and test set. Here we use XGBoost regressor and linear regression as machine learning models for comparison and then evaluate this models. The tool used is Google Colaboratory or Google Colab is a web based tool and a cloud-based service.

#### Colab

“Colab” short for Colaboratory, is a product from Google Research that runs completely in the cloud. Colab allows us to execute python code in the browser, and is mainly well suited for machine learning, data analysis and algorithms. On the subject of technical, Colab is a Jupyter notebook offers a hosted service which want no setup to use, at the same time as put forward loose get right of entry to to computing resources comprising GPUs.

Colab may be used perform the full strength of well-known Python libraries to investigate and visualize facts. With Colab you may import an image dataset, train an image classifier on it, and examine the model, all in only some lines of code. Colab notebooks execute code on Google

cloud servers, which means you have an advantage of Google hardware, as well as GPUs and TPUs, nevertheless of the power of your machine. All you need is a browser. A programmer can perform the using Google Colab. We can write and execute code in Python in colab.

- Document your code that supports mathematical equations
- Create, Upload, Share notebooks
- Import and save notebook from or to Google Drive
- Import or Publish notebooks from GitHub
- Import external datasets e.g. from Kaggle
- Integrate PyTorch, TensorFlow, Keras, OpenCV
- Free Cloud service with free

### V. IMPLEMENTATION

Here we are using 2 kind of algorithms on given data set in order make out a useful model for predicting the calories burned during the workout based on the workout duration, age, gender, height and weight of the person.

#### Algorithms

**RGBoost Regressor** - RGBoost Regressor is a regression algorithm that is an analysis which is a statistical technique to model the connection between a dependent (target) and independent (predictor) variables with one or more unbiased variables. In machine learning the XGBoost algorithm performs well since it has robust handling of many variety of data types, relationships, distributions, and the many hyperparameters that you can fine-tune. XGBoost regressor can be used for regression, classification for both binary and multiclass, and ranking problems.

**Linear Regression** – In Machine Learning, Linear Regression is a supervised learning algorithm. In this the linear equation specify the relationship amongst one or extra predictor variables and one outcome variable and combines a specific set of input values (x) to which is the predicted output for that set of input values (y). The equation can be written as  $Y = a + bX$ .

#### Steps:

- A. Collect Dataset
- B Data Pre-processing
- C Data Analysis
- D. Machine learning model
- E. Evaluation

**Data Source**

The repository that we used for dataset is Kaggle. There are two csv files which contains 15000 instances and 7 attributes. The data set from Kaggle repository contains attributes of each person’s details including their gender,age,workout duration, heart rate, body temperature, height and weight. This dataset is taken as the training data. And the second calories dataset contains target class which have the calories burned by corresponding person.

Table 1: Attributes and its values

Input Attributes	Function
gender	gender (male : 0,female : 1)
age	age mentioned in years
height	height of the person
weight	weight of the person
duration	The time taken to complete the exercising in minutes.
heart_rate	Average heart rate during the workout (more than normal rate 75 beats/min)
body_temp	Body temperature in the course of the workout(greater than 37 degree celsius)
calories	Total calories burned during the exercise.

There are two dataset csv files which should be uploaded to colab which is used for processing. We use data frames for analysis and processing. It obtain some statistical measures about the data.

```
[ ] calories_data.describe()
```

	User_ID	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
count	1.500000e+04	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000
mean	1.497739e+07	42.789800	174.465133	74.966867	15.530600	95.518533	40.025453	89.539533
std	2.872851e+06	16.980264	14.258114	15.035657	8.319203	9.583328	0.779230	62.456978
min	1.000116e+07	20.000000	123.000000	36.000000	1.000000	67.000000	37.100000	1.000000
25%	1.247419e+07	28.000000	164.000000	63.000000	8.000000	88.000000	39.600000	35.000000
50%	1.499720e+07	39.000000	175.000000	74.000000	16.000000	96.000000	40.200000	79.000000
75%	1.744928e+07	56.000000	185.000000	87.000000	23.000000	103.000000	40.600000	138.000000
max	1.999965e+07	79.000000	222.000000	132.000000	30.000000	128.000000	41.500000	314.000000

Figure 1- dataset

Here the mean value of age is 42.7 and we have the standard deviation and the percentile information. The body temperature is about 40. The body temperature will be more in those people who are doing workout. The crucial insights for this analysis is coronary heart rate and temperature. Subsequently we need to visualize the data via using a few plots and graphs.

**DOI: 10.5281/zenodo.6365018**

ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam



Figure 2- gender\_plot

we have almost good identical distribution of male and female data points.

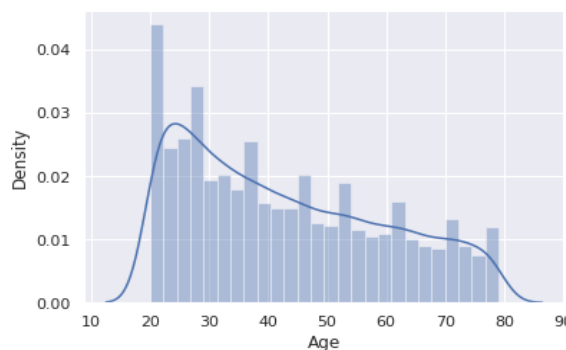


Figure 3 - age\_plot

For age, the more values in the range of 20 and 30 where we can see a peak in the curve means from 15000 instances more are present in this range. Then there is a decreasing means less people tend to do workout at older age.

Then we study the correlation the various records and there are two types of correlation, positive and negative correlation. The duration for which the person doing the workout is more then the number of calories the person burnt will also be more. So those values are immediately proportional which is inside the equal direction are undoubtedly correlated.

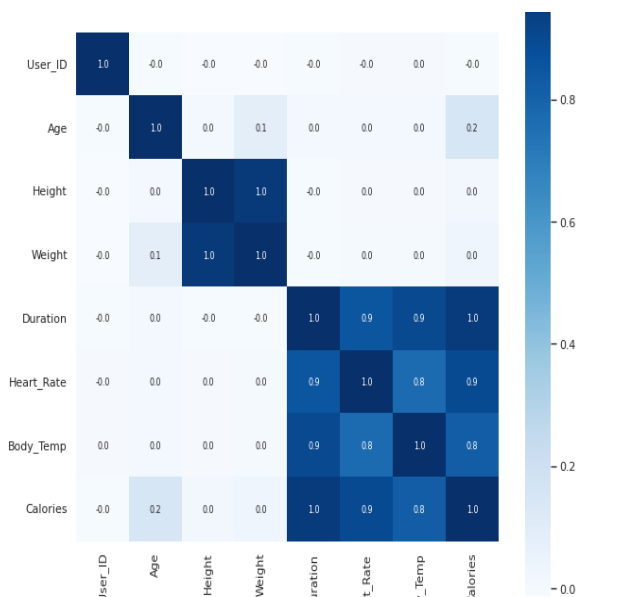


Figure 4- correlation\_heatmap

From this we get that the weight and weight are positively correlated and the duration, heart rate and body temperature are highly positively correlated with calories. Now split the data into training and test data by taking the two variables X and Y that is to separate for features and target.

```
[ ] X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
```

Then load the model XGB Regressor and evaluate the prediction on test data. The model goes through this test data and calories burned for the X\_test. And compare the values predicted by our model with original values.

For this we use the metrics-mean absolute blunders which tells what is the magnitude of mistakes the version is making metrics.mean\_absolute\_error(Y\_test, test\_data\_prediction)

The mean absolute error getting is 2.71.

```
input_data = (0,68,190.0,94.0,29.0,105.0,40.8)
input_data_as_numpy_array = np.asarray(input_data)
input_data_resaped = input_data_as_numpy_array.reshape(1,-1)
prediction = model.predict(input_data_resaped)
print(prediction)
[230.33057]
```

Figure 5 – XGBRegressor output

Next we do this in linear regression model.

- lin\_reg\_model = LinearRegression()

The mean absolute error getting is 8.38.

```
input_data = (0,68,190.0,94.0,29.0,105.0,40.8)
# change the input data to a numpy array
input_data_as_numpy_array= np.asarray(input_data)
# reshape the numpy array as we are predicting for only on instance
input_data_resaped = input_data_as_numpy_array.reshape(1,-1)
prediction = lin_reg_model.predict(input_data_resaped)
print(prediction)
[199.38038663]
```

Figure 6 – Linear regression output

## VI. RESULT

The analysis of this dataset was done to predict the calories burned depends on the duration of workout and also based on the gender, age, body temperature and heart rate at some stage in the exercise. By using these machine learning algorithms we are looking for a machine learning model with less mean absolute error, which gives more accurate results. By comparing the two models, XGB regressor and Linear regression we get that the XGB Regressor gives the more accurate results of the calories burned with a mean absolute error of 2.71 than the linear regression.

Machine Learning Model	Mean Absolute Error
XGB Regressor	2.71
Linear Regression	8.38

Table – 1

Machine Learning Model	Input data	Predicted Calorie result	Expected Calorie result
XGB Regressor	Male{0},68,190.0,94.0,29.0,105.0,40.8	230.33	231.0
Linear Regression	Male{0},68,190.0,94.0,29.0,105.0,40.8	199.380	231.0

Table - 2

## VII. CONCLUSION

From the analysis we met with a conclusion that the XGB Regressor has more accurate results than the Linear regression model. Mean absolute error imply absolute error ought to be as low as viable.. it is not anything but the difference between the actual and predicted values through the models. The mean absolute error value that is getting in XGB Regressor is 2.71 which is a good value. The error values is very less. Therefore we can conclude that the best model for the calorie burn prediction is XGBoost Regressor.

## VIII. REFERENCES

- [1] <https://www.kaggle.com/fmendes/fmendesdat263xdemos>
- [2] <https://machinelearningmastery.com/linear-regression-for-machine-learning/>
- [3] <https://machinelearningmastery.com/xgboost-for-regression/>
- [4] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5496172>
- [5] <https://devhadvani.github.io/calorie.html>