IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96125

Calorie Burn Predection using Machine Learning

Suvarna Shreyas Ratnakar[1], Vidya S[2]

[1] Student, Dept. of MCA, Bangalore Institute of Technology, Bengaluru, India

[2] Assistant Professor, Dept. of MCA, Bangalore Institute of Technology, Bengaluru, India.

Abstract: In today's world, people are having very tight schedules due to the changes in their lifestyles and work commitments. But it requires regular physical activity to stay fit and healthy. People do not concentrate on their food habits, leading to obesity. Obesity is becoming a major and common problem in today's lifestyle. This leads people to choose their diet and do an equal amount of exercise to stay fit and healthy. The main part here is people should have adequate knowledge about their calorie intake and burn, keeping a track of their calorie intake is easy as it's available on the product label or on the internet. Keeping track of calories burnt is a difficult part as there are very few devices for that. Calories burned by an individual are based on MET charts and formulas. The main agenda of this study is a prediction of the burnt calories with the help of an XG boost regression model as the ML (machine learning) algorithm to show accurate results. The model is fed with more than 15,000 data and its mean absolute error is 2.7 which will become better over time by feeding the XG boost regression model with more data.

I. INTRODUCTION

Most often, when individuals think of calories, they only think of food or weight reduction. A calorie, however, is often a measure of heat energy. Calories are the units of energy needed to elevate 1 gramme (g) of water by 1°C. The measurement may be used to assess a variety of energy-releasing systems unrelated to the human body. The amount of energy needed by the body to carry out a task is how many calories are considered from the perspective of the human body. There are calories in food. Each and every item has a distinct quantity of energy included in it since various foods have varying calorie counts.

The temperature of the body and the heartbeat will start rising up when we perform exercise or some heavy workout. The carbohydrates or carbs are broken down into glucose which is further converted/broken down into energy using O2(oxygen). The variables used here are the timescale the person is training, the average heart rate per minute, and the temperature. Then get more height, weight, gender, and age of the person to predict the tonnage of energy that the person burns. Parameters that can be considered for input are the duration of exercise, average heart rate per minute, temperature, height, weight, and gender. A machine learning XGBoost regressor algorithm is used to predict calories burned depending on exercise time, temperature, height, weight, and age.

II. LITERATURE SURVEY

Daniel Bubnis [6] The variety of burned energy in day-to-day life is directly related to weight maintenance, weight gain, or weight loss. People need to burn more calories than they consume, causing a calorie deficiency. But they want to know how many calories they burn every day. Most people think that calories are most effectively associated with food and weight loss. Calories are variously defined units of energy or heat. For men or women trying to gain, lose, or maintain weight, it is essential to know how many calories they are consuming each day.

Salvador Camacho [9] The global obesity crisis has been continuously increasing, and thus far no nation has been able to turn it around. The World Health Organization identifies an energy imbalance between calories ingested and calories expended as the root cause of obesity. But mounting data indicates that the idea of calorie imbalance might not be enough to control and stop the obesity pandemic.

To examine the calorie imbalance idea and its components as a weight-management tool as well as any potential drawbacks and implications for public health, with the goal of highlighting the need for an updated theory about the origins of obesity. This revision could better direct public health initiatives to control obesity by avoiding weight increase or encouraging weight reduction.

World Health Organization. (2011, October) Obesity Study. [Online] [10] Understanding the factors that affect calorie burning might help someone modify their diet or exercise routine to achieve the desired results. There is numerous studies in literature that use machine learning and data mining to diagnose these problems. When compared to today's study, some articles published two to three years ago have a lower accuracy for predicting calories burnt problems.

III. METHODOLOGY

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96125

In order to determine how many calories an individual would burn, this study is all about gathering the right data set to train our machine learning models. Pre-processing of the records is necessary before the statistics feeding operation. After that, data processing is completed, and the data is arranged as plots and graphs using several visualisation techniques. Here we use XG Boost regressor as a ML(machine learning) model for making comparison and then evaluating these models

Work Flow:

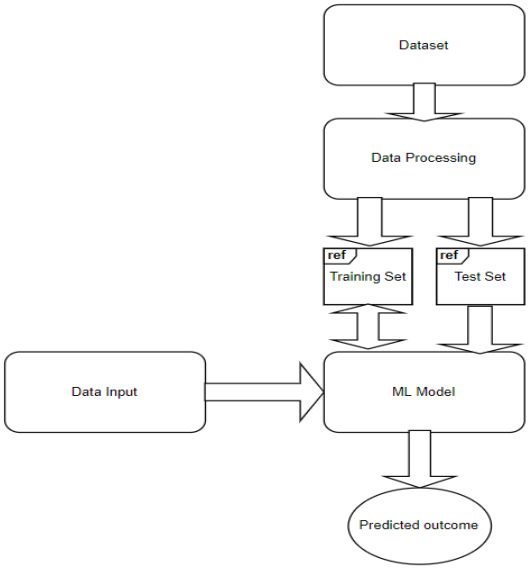


Figure 1 - Work Flow

Data source:

We use "Kaggle" as our dataset store. A total of 15000 instances and 7 attributes of data are present throughout 2 CSV files. The "Kaggle" repository's dataset comprises information about a variety of people, including their height, weight, gender, age, workout intensity, heart rate, and body temperature. The training data is taken from the “exercise.csv” and “calories.csv” datasets. Additionally, the user id-mapped target class from the second calorie dataset comprises the calories burned by the person in exercise dataset.

Table 1: Attributes and their values

Attribute	Function
Gender_individual	Gender (female : 1, male : 0)
Age_Individual	Age is mentioned in years
Height_Individual	Height of the person
Weight_Individual	Weight of the person
Heart_rate_Individual	Average heart rate of an individual during the workout(Normal heart rate 75 beats/min)
Body_temp_individual	Average body temperature captured in the course of entireworkout (greater than 37 degree Celsius)
Duration_individual	Duration of exercising in minutes.
Calories_individual	The total amount of calories burned while workout.

There is 2 datasets in CSV files format that is necessary to be uploaded in collab which is mainly used for data-

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96125

processing online. Data frames are been used for processing and analysis purposes. This has obtained some statistical measures of the data.

```
calories_data.head()
```

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
0	14733363	0	68	190.0	94.0	29.0	105.0	40.8	231.0
1	14861698	1	20	166.0	60.0	14.0	94.0	40.3	66.0
2	11179863	0	69	179.0	79.0	5.0	88.0	38.7	26.0
3	16180408	1	34	179.0	71.0	13.0	100.0	40.5	71.0
4	17771927	1	27	154.0	58.0	10.0	81.0	39.8	35.0

Figure 3 - data frame

The description of the above dataset can be seen in fig 4.

```
# get 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	1.497736e+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%	1.499728e+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 4 - Data Description

The count of the gender is equally distributed in the dataset which can be seen in the below figure (figure 3).

```
# plotting the gender column in count plot
sns.countplot(calories_data['Gender'],palette="Set1")
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorator
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f4303134110>
```

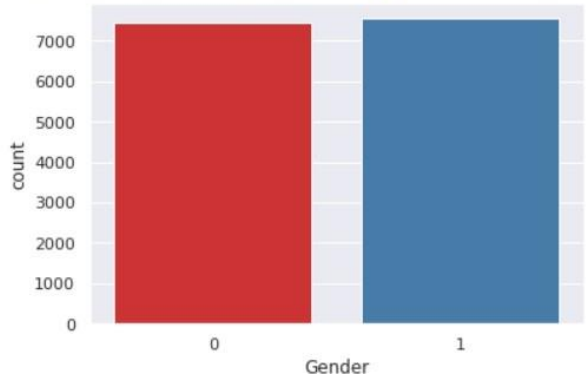


Figure 5 - Gender Distribution

Also, we have the mean value for age, height, and weight represented in the figure below.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96125

```
# finding the distribution of "Age" column
sns.distplot(calories_data['Age'], color = "R")
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2681: MatplotlibDeprecat
  color=hist_color, **hist_kws)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:177: MatplotlibDeprecati
  kws["color"] = to_rgba(color, alpha)
<matplotlib.axes._subplots.AxesSubplot at 0x7f43031cbe10>
```

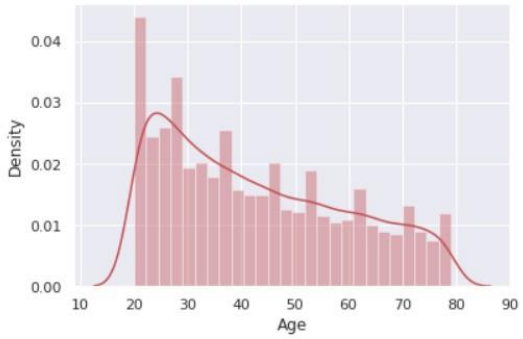
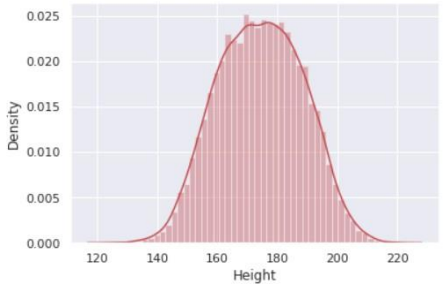


Figure 6 - Mean Age

For age, more values between the age group of 20 and 30 can be seen. There is a peek in the curve means which we had generated using 15000 instances. The decrease in the curve can be seen as people tend to not workout at an older age.

```
# finding the distribution of "Height" column
sns.distplot(calories_data['Height'], color = "R")
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributio
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f43032dd710>
```




```
# finding the distribution of "Weight" column
sns.distplot(calories_data['Weight'], color = "R")
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributi
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f4302dffe50>
```

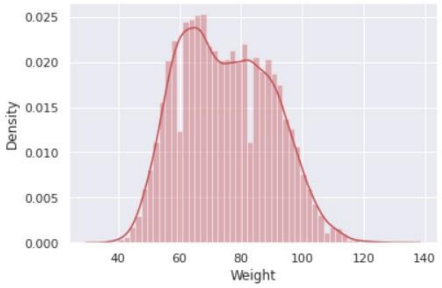


Figure 7 - Mean Height and Weight

The relationship between the different records is then examined. There are two forms of correlation: positive correlation and negative correlation. The quantity of calories burned will increase as exercise duration increases. These values are therefore proportionate, i.e., in the same direction, and unquestionably connected.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96125

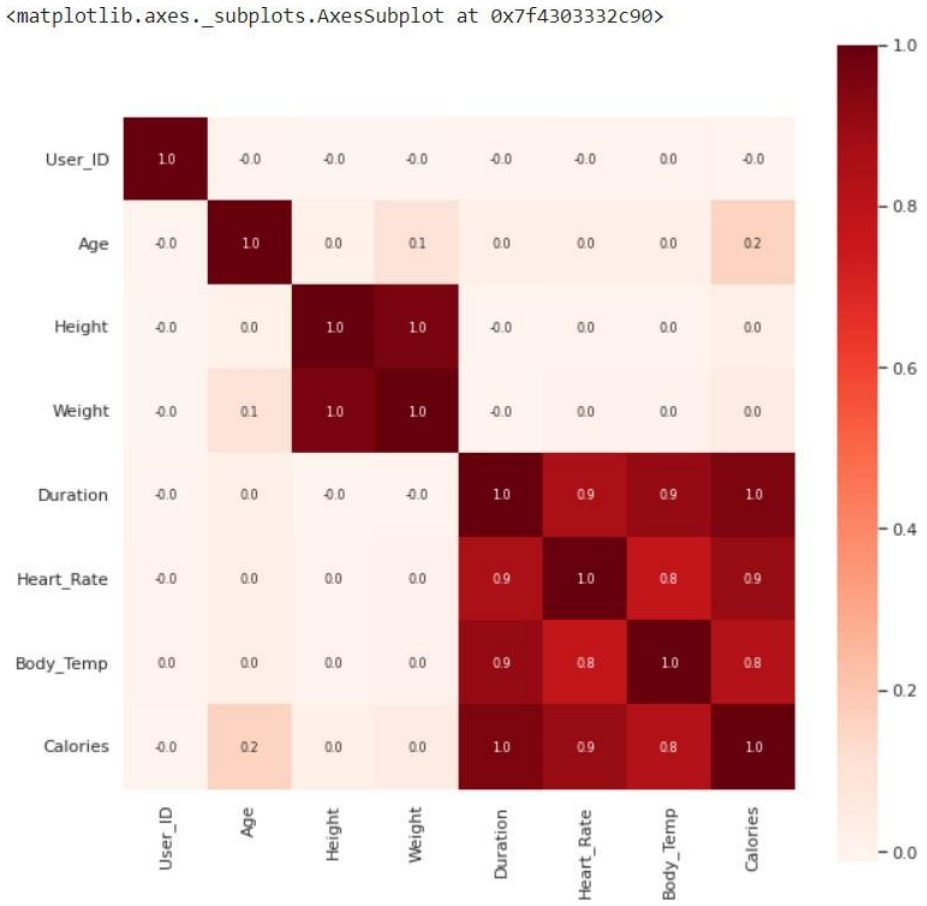


Figure 8 - Correlation of attribute

A. Collecting dataset

Data retrieval is the first and initial step. Kaggle is the data repository we use. It's loaded into the Collab program. The information gathered is both category and numerical.

B. Data Pre-processing

15000 instances and 7 attributes of data are contained in two csv files(“exercise.csv” and “calorie.csv”. Each person's attributes are included in the Kaggle data collection, including their height, weight, gender, age, workout duration, heart rate, and body temperature.

Data pre-processing is a crucial step in the machine learning process since the calibre of the data and the information that can be extracted from it directly affects how well our model can learn. It is crucial that we preprocess our data before supplying it to our model as aresult.

C. Data Analysis

Colab, the platform used for processing, requires the upload of two dataset csv files (“exercise.csv” and “calorie.csv”). The average body temperature is 40. Those who are exercising will have a higher body temperature. The coronary heart rate and temperature are the most important findings for this analysis. The data must then be visualized using a few charts and graphs. The two types of correlation—positive and negative correlation—are then studied between the various records. After that, load the XGB Regressor model and assessthe prediction using test data. This test data and calories burned for the X test are run through the model. Likewise, contrast our model's projected values with the original values.

D. Machine Learning model

This is the stage in which we apply the algorithm we've chosen (in this case, XGBoost regressor) to determine the mean absolute error. The XGB regression procedure is used, and the results are obtained. For this, we employ metrics that indicate the magnitude of errors the version is committing. The XGBoost regressor algorithm was proven to be an

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96125

effective and efficient method in predicting the calories burnt prediction.

E. Evaluation

This dataset was analyzed to make predictions about the number of calories burned depending on the length of the workout as well as on factors like age, gender, body temperature, and heart rate at various points during the exercise. We are searching for a machine learning model with a lower mean absolute error that produces more accurate outcomes utilizing these machine learning methods.

IV. RESULT

A. Dataset's first five rows:

Table view of the first 5 records in the dataset:

```
calories_data.head()
```

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
0	14733363	male	68	190.0	94.0	29.0	105.0	40.8	231.0
1	14861698	female	20	166.0	60.0	14.0	94.0	40.3	66.0
2	11179863	male	69	179.0	79.0	5.0	88.0	38.7	26.0
3	16180408	female	34	179.0	71.0	13.0	100.0	40.5	71.0
4	17771927	female	27	154.0	58.0	10.0	81.0	39.8	35.0

Figure 9 - data frame

B. Conversion of text data into numerical form:

```
[21] calories_data.replace({"Gender":{'male':0,'female':1}}, inplace=True)
```

```
[22] calories_data.head()
```

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
0	14733363	0	68	190.0	94.0	29.0	105.0	40.8	231.0
1	14861698	1	20	166.0	60.0	14.0	94.0	40.3	66.0
2	11179863	0	69	179.0	79.0	5.0	88.0	38.7	26.0
3	16180408	1	34	179.0	71.0	13.0	100.0	40.5	71.0
4	17771927	1	27	154.0	58.0	10.0	81.0	39.8	35.0

Figure 10 - Gender data conversion

C. Splitting of Data:

Splitting the data into training data and Test data

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

```
[ ] print(X.shape, X_train.shape, X_test.shape)

    (15000, 7) (12000, 7) (3000, 7)
```

Figure 11 - Splitting of data

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96125

D. Data Training in XG Boost Regressor

```
[ ] # loading the model
    model = XGBRegressor()

[ ] # training the model with X_train
    model.fit(X_train, Y_train)

    [16:36:32] WARNING: /workspace/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.
    XGBRegressor()
```

Figure 12 - Training of data

E. Mean absolute error:

Figure 13 - Absolute mean of data

```
[ ] test_data_prediction = model.predict(X_test)

[ ] print(test_data_prediction)

    [129.06204  223.79721   39.181965 ... 145.59767   22.53474   92.29064 ]
```

Mean Absolute Error

```
[ ] mae = metrics.mean_absolute_error(Y_test, test_data_prediction)

[ ] print("Mean Absolute Error = ", mae)

    Mean Absolute Error =  2.7159012502233186
```

V. CONCLUSION

We deduced from the analysis that the XGB Regressor produces more accurate findings. Mean absolute error suggests that absolute error should be as minimal as possible. It is nothing more than the discrepancy between values that were seen and those that were predicted by models. 2.71 is a good value for the mean absolute value that the XGBRegressor gives us. The mistake rates are quite low. Therefore, we can say that XG Boost Regressor is the best model for predicting calorie burn. The flexibility of the suggested technique can also be improved with variations.

In this study, we have concentrated on the seven primary factors that influence how many calories our body burns, but there are other factors that also play a role. It's also crucial to understand how many calories we are consuming if we want to stay healthy and fit. Additionally, ML may be used to construct this (machine learning). A UI (user interface) is also required so that users may input their values and obtain results that show how many calories they have burned. Additionally, we are able to create a completely functional app with all of these features and our recommended diet and exercise regimen.

REFERENCES

[1] Smola, A., & Vishwanathan, S. V. N. (2008). Introduction to machine learning. *Cambridge University*.
[2] MacKay, D.J, & Mac Kay, D. J. (2003). *Information theory, inference and learning algorithms*.Cambridge University press.
[3] Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*
[4] https://www.kaggle.com/fmendes/fmendesdat263xdemos
[5] https://machinelearningmastery.com/xgboost-for- regression/
[6] https://www.medicalnewstoday.com/articles/319731#factors-influencing-daily-calorie-burn-and-weight-loss
[7] https://zenodo.org/record/6365018
[8] https://devhadvani.github.io/calorie.htm
[9] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5496172
[10] https://www.jetpac.com/
[11] World Health Organization. (2011, October) Obesity Study. [Online]http://www.who.int/mediacentre/factsheets /fs311/en/index.html