Calorie Burn Prediction Analysis

Using XGBoost Regressor

Team:

Likhit Lingudu
Divija Surepally
Jayakanth Srinivasan
Sai Hruthik Reddy Katipally

Abstract:

The project idea is to predict the Calorie Burnt during the exercise using the XGBoost Regressor Algorithm. In this project, we build the Machine Learning system to predict the calories burnt during exercise. Nowadays, people are keenly interested in the workout plans, how many calories they've burnt in the workout period, etc. People who are on a diet/exercise plan can use of this data to modify their habits and maintain ideal health. This is the motivation behind this project.

Introduction:

When people work out, body temperature and heartbeat rise. Variables such as average heartbeat, body temperature, and duration of the workout are considered. Then we consider the height, weight, age, and gender of the person and predict how many calories have been burnt during the workout period. The XGBoost Regressor algorithm depends on the following data to predict the calories burnt: height, weight, age, gender, heartbeat rate, body temperature, and time of workout duration.

Description:

The energy burned each day is connected to weight loss, weight gain, or weight maintenance. To shed weight, a person must burn greater calories than they take in. To do that, they want to recognize what number of calories they burn each day. Understanding what factors will help in burning the calorie will help the people a great deal in regulating a diet and proper exercise pattern. These factors are in our control, and some are not. Factors include age, gender, body composition etc.

Implementing XGBoost Regressor:

XGBoost Regression is a statistical technique to model the connection between dependent (target) and independent (predictor) variables with one or more unbiased variables. In Machine Learning, XGBoost algorithm performs well since it has robust handling of many varieties of data types, relationships, distributions, and several hyperparameters that you can fine-tune. XGBoost regressor can be applied for regression, classification for both binary and multiclass, and ranking problems.

Regression Analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables.

Dependent Variable: factor that we're trying to predict.
Independent Variables: factors that have an impact on the dependent variable.

Steps Followed:

1. Collect Dataset
2. Data Preprocessing
3. Data Analysis
4. Train-Test-Split
5. Regressor Model
6. Evaluation

Data Source:

The Data set on which we use for the prediction is available on Kaggle. There are .csv files that contain 15000 instances and 7 attributes. The data set from the 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.

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 during the workout (greater than 37 degree Celsius)
calories	Total calories burned during the exercise.

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

Splitting the data into training and test data:

- Using **train_test_split**, the dataset is divided into training sets and Test sets.
- Training set - a subset of Data set to train a model.
- Test set - a subset of Data set to test the trained model.
- **X** contains the original amount of data which is 15000
- **xTrain** contains 80% of the data which is 12000
- **xTest** contains 20% of the data which is 3000

Prediction:

- A tree-based algorithm called XGBoost Regressor will be used for the prediction.
- Higher accuracy and better performance than the Random Forest algorithm.
- Dependent variable: number of Calories burned.
- Independent variables: gender, age, height, weight, duration of exercise, heart rate, and body temperature.

Mean Absolute Error: 1.48

Prediction Mean: 89.34

Results:

The analysis of this dataset was done to predict the calories burned depending on the duration of the workout, and also based on the gender, age, body temperature, and heart rate at some stage in the exercise. By using the XGBoost Regression algorithm we are looking for a machine learning model with less mean absolute error, which gives more accurate results.

Conclusion:

From the analysis, we conclude that the XGB Regressor has more accurate results. The difference between the actual and predicted values through the XGBoost regressor model is low. We can conclude XGBoost Regressor Model is very much suited for the Calorie Burn Prediction.

GitHub: https://github.com/JayGitProfile/Calorie-prediction-xgboost-regressor/blob/main/DAMT%20project.ipynb

References:

- https://www.kaggle.com/fmendes/fmendesdat263xdemos
- https://machinelearningmastery.com/xgboost-for-regression/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5496172/
- https://devhadvani.github.io/calorie.html