Calorie Burn Prediction Through Modelling

By Mitra Zamani

16 July 2023

Summary:

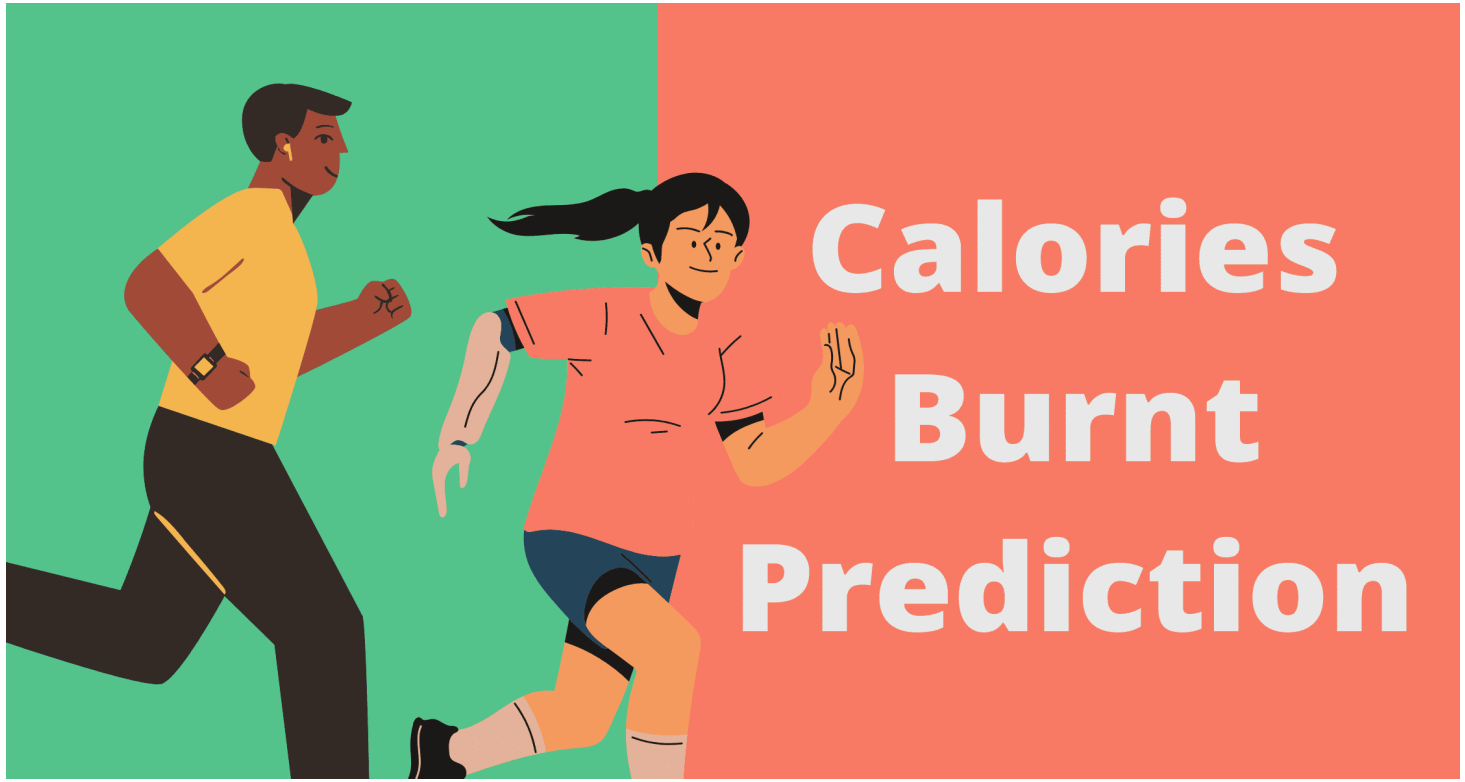
- Most often, when individuals think of calories, they only think of food or weight reduction. A calorie, however, is often a measure of heat energy. Calories are the units of energy needed to elevate 1 gramme (g) of water by 1°C. The measurement may be used to assess a variety of energy-releasing systems 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. linear regression is used to predict calories burned depending on exercise time, temperature, height, weight, and age.

Outline:

- Business Problem
- Data and Methods
- Results
- Conclusions

Business Problem

predict which biological measures are more important and effective to burn more calories during a workout.



Data

- This dataset contains tabular data for gender, age, height, weight, duration of exercise done, heart beat rate, body temperature, and calories burnt from 15000 samples. Hence it is possible to develop a regression model for calories burnt by using this available data.

Methods

- This project will employ the ‘OSEMN’ data science process to source the data, perform exploratory data analysis, clean and prepare the data for model, fit regression models, interpret results that will be used to predict the number of calories a person has burnt during a workout.

Results

After cleaning Data and some changes on categorical data, I created model. The model's performance was good with an r-squared of 0.947 , explaining 95% of the variance

Out[27]:

OLS Regression Results

Dep. Variable:	Calories	**R-squared:**	0.947
Model:	OLS	**Adj. R-squared:**	0.947
Method:	Least Squares	**F-statistic:**	6.640e+04
Date:	Fri, 21 Jul 2023	**Prob (F-statistic):**	0.00
Time:	22:58:17	**Log-Likelihood:**	-61332.
No. Observations:	15000	**AIC:**	1.227e+05
Df Residuals:	14995	**BIC:**	1.227e+05
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	472.9347	13.771	34.344	0.000	445.943	499.927
Gender	-2.1630	0.236	-9.171	0.000	-2.625	-1.701
Duration	6.6471	0.040	165.209	0.000	6.568	6.726
Heart_Rate	1.9877	0.024	84.339	0.000	1.942	2.034
Body_Temp	-16.8744	0.353	-47.870	0.000	-17.565	-16.183

Omnibus:	2720.259	**Durbin-Watson:**	2.011
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	8318.374
Skew:	0.940	**Prob(JB):**	0.00
Kurtosis:	6.127	**Cond. No.**	1.23e+04

Train-test-split and cross-validation:

Then I did Model validations (Train/Test Split) and Calculated the Mean Squared Error (MSE). Train MSE is lower than Test MSE, so our model is overfitting. Overfitting occurs when the model cannot generalize and fits too closely to the training dataset instead. this is a sign that the model doesn't generalize well to future cases. After that I did cross validation, I compared the result of that with train-test-split. So the train-test split result is about 195 whereas the average cross-validation result is about 192. Because this is an error-based metric, a higher value is worse, so this means that the cross-validation result is "better" (more optimistic).

Conclusions

- We deduced from the analysis that the Linear regression model produces more accurate findings. The R-squared is 0.94 and The mistake rates are quite low. Therefore, we can say that linear Regression is the best model for predicting calorie burn. Our model was overfitting so for detecting overfitting I did cross validation. 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. The most effective biological measures for burning calories are duration of doing exercise, heart_rate and body temperature. so I recommend to you if you want to burn more calories increase the duration of your exercise.

- **Thank you!**
- Email: zmitra15@gmail.com
- GitHub: @MitraZamani