Burned Calories Prediction using Supervised Machine Learning: Regression Algorithm

Marte Nipas
Computer Engineering Department
Technological Institute of the Philippines
Manila, Philippines
marte.nipas@tip.edu.ph

Aimee G. Acoba
Computer Engineering Department
Technological Institute of the Philippines
Manila, Philippines
aacoba10@gmail.com

Jennalyn N. Mindoro
Computer Engineering Department
Technological Institute of the Philippines
Manila, Philippines
jnicolas.cpe@tip.edu.ph

Mon Arjay F. Malbog
Computer Engineering Department
Technological Institute of the Philippines
Manila, Philippines
malbog.monarjay@gmail.com

Julie Ann B. Susa
Computer Engineering Department
Technological Institute of the Philippines
Manila, Philippines
jannsusa@gmail.com

Joshua S. Gulmatico
Computer Engineering Department
Technological Institute of the Philippines
Manila, Philippines
engr.gulmatico@gmail.com

2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T) | 978-1-6654-5858-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICPC2T53885.2022.9776710

***Abstract*— Regular physical activities are essential to staying healthy and fit. The estimation of calories burned by individuals is based on a formula and MET charts. This study aims to predict the calories burned using a regression model as one of the machine learning algorithms to give more accurate results. Data preparation, cleaning, and analysis are the primary steps before they can be fed to the regression models. Model training and testing using K-fold validation were done to determine the best model for the study. The performance and prediction accuracy of regression models were evaluated based on the result of model testing after ten (10) iterations. The average accuracy was computed and the result shows that Random Forest regression is the best model for the study with an accuracy of 95.77%. It is very important to visualize and study the relationships of the variables in the data because it may affect the performance of the algorithm in predicting the value of the target variable. The Random Forest regression model was able to predict the calories burned with a high accuracy rate.**

Keywords*— *calories, regression model, prediction, machine learning

I. Introduction

A calorie is a unit of energy that comes from the foods and drinks consumed by a person and the energy used in physical activities [4]. The number of calories present in food varies and depends on the energy it can provide. Therefore, proper consumption of calories is important because it may lead to obesity, diabetes, and other related health problems [8].

In today's modern living, regular physical activity is important to become fit, in good shape, and to maintain a healthy body or to lose weight. Physical activities like running, walking, bicycling, swimming, exercising, or doing regular daily tasks burn calories [5]. The calories burned depends on many factors such as weight, gender, age, height, metabolism, and the type of activity or exercise done.

Measurement of exact calories burned can be difficult. The use activity tracker app can estimate the calorie burned, but the accuracy varies from product to product. Heart rate is one of the best ways to measure calorie burn, but it differs significantly between individuals based on fitness, age, and genetics [3]. Metabolic Equivalent of Task (MET) is another way to estimate calorie burned developed by researchers and used by medical community. MET is the ratio of active metabolic rate to resting metabolic rate [12]. Metabolic rate is the rate of energy consumed per unit of time which is based on the intensity of the activity or exercise.

The study aims to propose a solution in predicting the calories burned using machine learning algorithms. The learning algorithms considered are Linear Regression, Ridge Regression, and Random Forest Regression. The goal of this study is to evaluate which algorithms would be the best in predicting calories burned by an individual based on weight, gender, age, height, duration of the activity, heart rate, and body temperature. The model produced by the study can be integrated or used with the existing technologies to have a better estimate of calories burned by individuals after some physical activities.

II. Literature Review

A. Calorie Expenditures

The heart rate of an individual signifies the condition of the body, and heart rate can be measured through pulses. In the study of [3], they developed a prototype for calculating the calories burned through heart rate. The prototype used optical technology and photosensors to measure the heart rate. The controller then detects and counts the pulses. The measured value will be the basis of burned calories after exercising.

In a mobile application created by [11], it is used to increase the physical activity of the smartphone users and to encourage them to exercise together. To measure the calorie burned, it uses the accelerometer sensor [6] built in the phone to detect the movement of the players and a wearable polar sensor to record the heart rate. A similar approach was made by [1][2][9], but it uses the smartphone's Global Position System (GPS) to track the movement of the players. The user must move so that the character in the virtual world will also move.

The prototype of [13] uses an accelerometer and microcontroller to measure the calories by calculating the distance traveled by an individual through jogging, walking, or running and applying a mathematical formula to compute the total calories burned. On the other hand, [14] used a gyroscope and wemos module and sent the acquired data to a mobile application for processing to determine the number of footsteps and calories burned in a single activity.

The proposed algorithm of [12] can calculate the calories expenditures based on predicted heat strain model(PHSM), which estimates the sweat rate and body temperature in response to a particular environment. In the study, heat balance equation was considered to estimate the calorie expenditure.

B. *Regression Models*

Machine Learning uses different algorithms to learn the relationships in the data make predictions based on patterns identified from the dataset. Regression is one of the machine learning techniques where the model can predict the output based on the given variables.

Linear regression model is used to identify the linear relationship of the input variables and output variables [7]. Random Forest regression uses ensemble learning techniques for regression. Ensemble learning is a method that combines predictions from different machine learning algorithms to have a higher accurate prediction than using a single model. Random Forest uses trees instead of the equation to complete the process of regression. It constructs several decision trees during training time and outputs the mean of the classes to predict all the trees [10]. Ridge regression is applied to data with multi-collinearity.

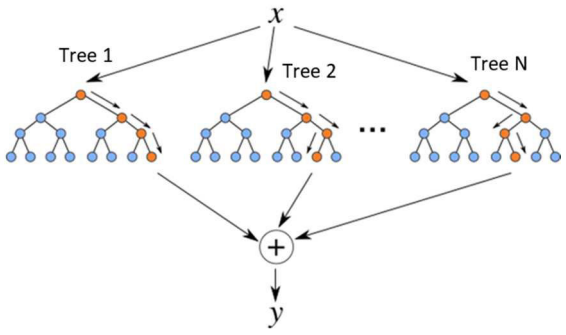

Fig. 1 Random Forest Regression Model

III. METHODOLOGY

A. *Data Collection and Preparation*

The data that was used in this study was taken from the Kaggle website. Table I shows the dataset specification that contains 15000 observations and nine variables. The Raw data gathered has eight (8) numeric variables and one (1) categorical variable. The data has no duplicate rows and no missing cells.

TABLE I. DATASET SPECIFICATION

Parameter	Value
Numeric variables	8
Categorical variable	1
Number of observations	15000
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.1MB
Average record size in memory	80.0B

B. *Data Analysis*

The variables in the dataset must be analyzed first in order to determine their relationship to the target variable, which is the calorie burned. The variables heart rate, duration of exercise, body temperature is highly correlated with the dependent variable calorie, followed by the height and weight.

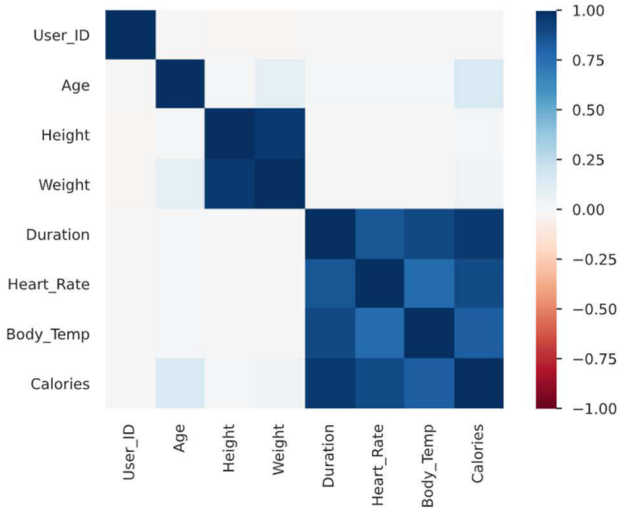

Fig. 2 Pearson's Correlation between Variables

Fig 3 shows the distribution of calories used for this study. From the distribution plot, it can infer that it doesn't follow the pattern of normal distribution.

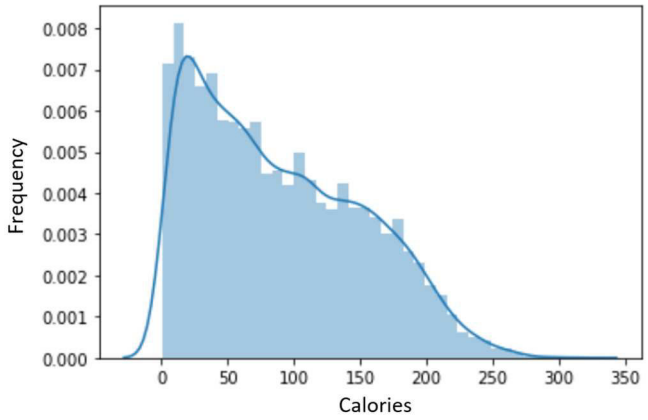

Fig. 3 Calorie Distribution

C. *Feature Extraction*

The data used in this study contains eight (8) features which consist of ID, Age, Height, Weight, Duration, Heart Rate, Body Temperature, and Gender. Based on the data analysis, the feature ID was not included because it doesn't have any impact in predicting the calorie burned of a person.

D. *Model Training and Prediction*

The study considered three regression models, namely linear regression, ridge regression, and random forest regression. The three models were trained with the dataset using the program developed in python programming. To test the models, K-fold cross-validation with ten iterations was used to reduce the chances of overfitting and to improve the prediction accuracy.

E. Model Assessment and Selection

Based on the result of testing, the models will be evaluated using the scores got by each model in 10 iterations of K-fold validation. The one with the highest average score in predicting the value of the target variable will be the basis in selecting the best model for the study. Also, prediction performance of each model was also considered to support the basis in selecting the best model. These performance measures are Mean Square Error(MSE), Root Mean Square Error(RMSE), and Mean Absolute Error(MAE). The model with the lowest prediction errors can be considered the best model.

IV. Result and Discussion

Table II shows the prediction performance of the three models. It shows that random forest has the lowest prediction errors and performs well compared to the other two models.

TABLE II. Model Performance of Regression Model

Metrics	Linear	Ridge	Random Forest
MSE	123.97	125.42	8.13
RMSE	11.13	11.20	2.85
MAE	8.23	8.30	1.81

Table III shows the K-fold validation results, and to determine the best model, compute the average prediction result of each model.

TABLE III. K-fold Validation Results

Iteration	Linear	Ridge	Random Forest
1	0.92704	0.926758	0.957204
2	0.929702	0.929207	0.957918
3	0.928701	0.927954	0.957368
4	0.928419	0.927704	0.957537
5	0.92942	0.928509	0.958222
6	0.927602	0.926676	0.957631
7	0.929818	0.929614	0.957539
8	0.929463	0.92904	0.957892
9	0.92779	0.927085	0.95767
10	0.930341	0.929758	0.957645

By taking the average, it can infer that Linear regression has an accuracy of 92.88%, Ridge regression is 92.82%, and Random Forest regression is 95.77%. This shows that Random Forest Regression is the best model for this study because of its capability to discover complex behaviors in the data being understudy. Table IV shows the result of sample data prediction using Random Forest regression. It was able to predict 14 correct predictions out of 15 with an accuracy of 93.33%.

TABLE IV. Random Forest Prediction Results

Test No.	Predicted Calorie	Actual Calorie	Interpretation
1	35	35	Correct
2	58	58	Correct
3	136	136	Correct
4	135	135	Correct
5	158	158	Correct
6	90	90	Correct
7	94	94	Correct
8	118	118	Correct
9	72	72	Correct
10	134	136	Incorrect
11	75	75	Correct
12	79	79	Correct
13	111	111	Correct
14	94	94	Correct
15	29	29	Correct

V. Conclusion

Regression models are one of the Machine Learning algorithms used to make a prediction. In this study it was able to predict the calories burned by an individual using a given dependent variable. After the identification of the problem, cleaning and preparation the data are needed before it can be feed on the algorithm. This study needs to be analyzed to determine the relationship present between each variable. It is important to visualize and understand the relationships between variables to look for problems like multi-collinearity and other issues. Because those problems have an impact on the algorithm that will be used. Plotting the samples in a 2-d plane can give us insights. In this study the researchers saw that the target variable has a visible correlation between heart rate, exercise duration, and body temperature. Lastly, the training and selection model is important and must be carefully conducted. Choosing the best fit model will make the prediction more accurate and reliable. The researchers achieve 95.77% accuracy using the random forest regression algorithm with the following hyper parameters: estimator is set to 100, maximum depth is none, minimum samples split is 2, minimum samples leaf is 1, and maximum features is auto.

As mentioned, the researchers did not modify any parameters in the regression models. For the researchers who might reference this study, it is recommended to alter some of the parameters to see if there are significant changes to the score. We also recommend trying other existing regression techniques that might fit better than what is used in this study.

Acknowledgement

The authors would like to show appreciation to the Technological Institute of the Philippines High-Performance Computing Laboratory for letting the researchers use their computing facilities.

References

[1] A. Koivisto, S. Merilampi, and K. Kiili. Mobile exergames for preventing diseases related to childhood obesity. In Proceedings of the 4th International Symposium on Applied in Sciences Biomedical and Communication Technologies, Barcelona, Spain, October 2011. 1

[2] C. G. Wylie. Mobile persuasive exergaming. In Proceedings of the 2009 International IEEE Consumer Electronics Society's Games Innovations Conference, Lancaster, UK, August 2009. 1

[3] G. K. Reddy and K. L. Achari, "A non invasive method for calculating calories burned during exercise using heartbeat," 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), 2015, pp. 1-5, doi: 10.1109/ISCO.2015.7282249

[4] S. Kasim and F. A. Zakaria, "Daily Calorie Manager for basic daily use," Third International Conference on Innovative Computing Technology (INTECH 2013), 2013, pp. 437-442, doi: 10.1109/INTECH.2013.6653675

[5] H. Rahaman and V. Dyo, "Counting calories without wearables: Device-free Human Energy Expenditure Estimation," 2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2020, pp. 1-6, doi: 10.1109/WiMob50308.2020.9253424.

[6] Jee Hyun Choi, Jeongwhan Lee, Hyun Tai Hwang, Jong Pal Kim, Jae Chan Park and Kunsoo Shin, "Estimation of Activity Energy Expenditure: Accelerometer Approach," 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005, pp. 3830-3833, doi: 10.1109/IEMBS.2005.1615295.

[7] M. S. Acharya, A. Armaan and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," 2019

International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862140.

[8] N. Komiya et al., "Novel Application of 3D Range Image Sensor to Caloric Expenditure Estimation based on Human Body Measurement," 2018 12th International Conference on Sensing Technology (ICST), 2018, pp. 371-374, doi: 10.1109/ICSensT.2018.8603651.

[9] P. Buddharaju and N. S. C. P. Pamidi, "Mobile Exergames - Burn Calories While Playing Games on a Smartphone," 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 50-51, doi: 10.1109/CVPRW.2013.13.

[10] P. Dong, H. Peng, X. Cheng, Y. Xing, X. Zhou and D. Huang, "A Random Forest Regression Model for Predicting Residual Stresses and Cutting Forces Introduced by Turning IN718 Alloy," 2019 IEEE International Conference on Computation, Communication and Engineering (ICCCE), 2019, pp. 5-8, doi: 10.1109/ICCCE48422.2019.9010767.

[11] S. Reiaz et al., "CalorieKiller: Burning Calories using Mobile Exergame with Wearables," 2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH), 2019, pp. 1-5, doi: 10.1109/SeGAH.2019.8882453.

[12] S. Yang, Y. Yeh, J. J. Ladasky and D. J. Schmidt, "Algorithm to calculate human calorie expenditure based on a predicted heat strain model," 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 2016, pp. 545-548, doi: 10.1109/BHI.2016.7455955.

[13] R. P. Sai, S. Bapanapalle, K. Praveen and M. P. Sunil, "Pedometer and calorie calculator for fitness tracking using MEMS digital accelerometer," 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-6, doi: 10.1109/INVENTIVE.2016.7823237.

[14] W. A. Kusuma, Z. Sari, H. Wibowo, S. Norhabibah, S. N. Ubay and D. A. Fitriani, "Monitoring Walking Devices For Calorie Balance In Patients With Medical Rehabilitation Needs," 2018 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2018, pp. 460-463, doi: 10.1109/EECSI.2018.8752761.