A Comprehensive Framework for Caloric Expenditure Estimation Utilizing Supervised Learning Techniques and Regression-Based Algorithms

Dharmaiah Devarapalli, Yannam Satya Amrutha, Bala Satya Sri Pasupuleti, Srujana Maddula, Hema Sri Puppala

***Abstract:** With the increasing importance of health and well-being in today's culture, exercise is becoming a significant element of daily activities. But often, individuals focus more on the outcomes of their efforts—such as how many calories they burn—than on the processes that produce them. This study presents the development of a prediction model that is integrated into a web application to determine an individual's caloric intake while engaging in physical exercise. The program examines key factors that significantly affect calorie burn using machine learning approaches, providing users with information on how effective their workouts are. To improve the predicted accuracy of the model, domain-specific parameters related to caloric expenditure were analyzed in this study. Heart rate, exercise duration, body temperature, height, and weight are among the factors selected for the model. Because it indicates the body's oxygen demand, which is a crucial component of the metabolic processes involved in producing energy from carbohydrates during physical exercise, heart rate is very important. Heart rate fluctuation is a useful predictor since it is correlated with the degree of exercise. The length of the exercise is also important because longer workouts tend to burn more calories. To account for individual physiological variations that impact energy consumption, body temperature, height, and weight were also taken into consideration. A dataset that recorded these characteristics during a variety of physical activities was used to train the model using supervised learning techniques. Accuracy, mean squared error and R-squared values were among the performance evaluation metrics used to confirm the model's ability to accurately estimate caloric expenditure. This research adds a useful application for users who want to monitor and enhance their physical health by giving them estimations that are customized to their unique qualities.*

***Keywords:** Supervised Learning; Regression Algorithm; Predictions; Calories.*

Manuscript received on 12 October 2024 | Revised Manuscript received on 26 October 2024 | Manuscript Accepted on 15 November 2024 | Manuscript published on 30 November 2024.
*****Correspondence Author(s)

Dr. Devarapalli Dharamaiah*, Department of Computer Science and Engineering, Koreru Lakshmaiah Educational Foundation, Vaddeswaram (Andhra Pradesh), India. Email ID: devarapalli.dharma@gmail.com, ORCID ID: 0000-0002-5804-5880

Ms. Yannam Satya Amrutha, Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram (Andhra Pradesh), India. Email ID: ysatyaamrutha@gmail.com, ORCID ID: 0000-0001-9888-7751

Ms. Bala Satya Sri Pasupuleti, Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram (Andhra Pradesh), India. Email ID: satyasri.10501@gmail.com, ORCID ID: 0000-0002-0768-8843

Ms. Srujana Maddula, Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram (Andhra Pradesh), India. Email ID: srujanamaddula05@gmail.com, ORCID ID: 0000-0002-2335-6232

Ms. Hema Sri Puppala, Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram (Andhra Pradesh), India. Email ID: puppalahemasri555@gmail.com, ORCID ID: 0000-0002-4379-7756

I. INTRODUCTION

Calories contained within foodstuffs serve to furnish energy in the form of thermal energy to our physiological systems, thereby enabling their operational functionality. This underscores the necessity for the ingestion of a specified quantity of calories to sustain life. Conversely, excessive caloric intake poses the hazard of weight gain. Consequently, there exists an imperative to expend calories. To achieve caloric expenditure, individuals engage in physical exercises and various activities aimed at caloric burning. To ascertain the quantity of calories expended, a caloric expenditure prediction system employs certain data to forecast caloric burn. Utilizing this software, one can estimate the caloric expenditure over a designated timeframe. This application incorporates various parameters, including age, sex, duration of exercise, and other factors, subsequently calculating the projected caloric burn. Users will also gain access to comparable outcomes and general information corresponding to the parameter values input into the application.

II. METHODOLOGY

A. Regression Algorithm

Regression analysis constitutes a method for elucidating the relationship between independent characteristics or variables and a dependent attribute or outcome. It represents a predictive modelling technique employed in machine learning, wherein an algorithm is utilized to anticipate continuous outcomes. One of the predominant applications of machine learning models, particularly in the realm of supervised machine learning, is the resolution of regression problems [1]. The association between independent and dependent variables is imparted to algorithms, which can then be deployed to project outcomes of novel, ambiguous incoming data or to address data deficiencies [1][8].

Rahmani and Mirmahaleh (2021) offer a comprehensive assessment of the literature that assesses several COVID-19 prevention and treatment strategies, emphasizing important factors that have an impact on health outcomes [9].

An interpretable decision tree is suggested by Sagi and Rokach (2021) as a way to simulate XGBoost, improving model transparency without sacrificing prediction accuracy [10]. Bentéjac, Csörgő, and Martínez-Muñoz (2020) compare .gradient boosting methods, assessing how well they work and how well they apply to different datasets [12][19][20][21][22][23].

Any forecasting or predictive model must include regression analysis, which is a prominent technique in machine learning-powered predictive analytics [14]. Regression is a common use for supervised machine learning models, along with classification [15]. Labelled input and output training data were required for this approach of training models. To comprehend the link between features and outcome variables, machine learning regression models need precise, labelled training data [1][8]. Regression may be found in many different machine-learning applications since it is a crucial part of predictive modelling [1][8].

Srikanth, Anusha, and Devarapalli (2015) describe a computational intelligence method that improves medical diagnosis efficacy by using decision tree algorithms [17].

Dash and Liu (1997) examine feature selection strategies for classification, emphasizing how they might increase model precision and effectiveness in intelligent data processing [16].

B. Ensemble Regression

Any forecasting or predictive model necessitates the incorporation of regression analysis, a prominent method within the domain of machine learning-powered predictive analytics. Regression is frequently employed alongside classification in supervised machine-learning models. This model training methodology requires labelled input and output training data. To fully comprehend the relationship between features and outcome variables, machine learning regression models necessitate precise, labelled training data. Regression is ubiquitous across various machine learning applications, given its integral role in predictive modelling. A method known as gradient boosting can be employed to address issues about classification or regression predictive modelling. Ensembles are constructed using decision tree models, wherein trees are sequentially added to the ensemble and calibrated to rectify the prediction errors introduced by preceding models. A variant of the ensemble machine learning model that corresponds to this methodology is referred to as boosting [2].

Model fitting may be conducted utilizing any differentiable loss function in conjunction with the gradient descent optimization technique. This methodology is termed "gradient boosting" due to the reduction of the loss gradient during the model fitting process, akin to the functioning of a neural network.

C. Regression using a Random Forest

The bagging technique is applied to multiple decision trees as part of an ensemble strategy known as Random Forest [5]. The fundamental concept revolves around amalgamating numerous decision trees to ascertain the outcome, rather than relying solely on a singular decision tree [4].

D. XGBoost Regressor

A gradient boosting framework is used by the ensemble machine learning technique known as XG boost, which is based on decision trees. Extreme Gradient boosting, often known as XG boosting, is the act of "boosting" or augmenting one weak model by merging it with several additional weak models to produce a model that is overall strong [3].

E. Decision Tree Regressor

Decision tree regression examines an object's features and trains a model in the shape of a tree to forecast future data and generate useful continuous insights. Using a tree structure that mimics a flowchart, the Decision Tree decision-making tool predicts the outcomes [6].

F. KNN Regressor

The KNN method uses "feature similarity" to forecast the values of fresh data points [15]. Thus, a value is determined for the new point based on how appropriate it is. To the training set's points, that is. Mean is used in KNN to compute the distance between data points [7].

III. DETAILS OF THE DATA SET

A. Data Collection

The practice of gathering information from numerous sources is known as data collection. For our project, we are compiling a list of various people's contact information. In this example, we're using a dataset with information on 1500 people, which is a collection of data. The attributes present in the dataset are user ID, gender, Age, Height(m), Weight(kg), Exercise duration(min), Heart rate, Body temperature and calories burnt.

B. Pre–Processing Step

The Data must be pre-processed before it can be trained and tested. There are various steps in data processing. Data visualization was applied during preprocessing to understand how the data is distributed, the correlation between calories and other attributes and also to get insight into the factors that impact calorie attributes. Data cleansing, integration, reduction, and transformation are a few examples of data preprocessing stages [13]. Since our data has already been cleaned, there is no need for data cleaning and integration; we simply converted the categorical values of gender into numerical values.

C. Separating the Training and Test Data from the Dataset

The training and testing portions of the data set are separated to determine how accurately the machine learning models will predict. The training set should receive 80% of the dataset, while the test set should receive the remaining 20%, as a rule. To tell whether the algorithm learned to generalize from the training set or simply memorized the data, it is critical to note that no data from the training set appears in the test set data.

Training Dataset: This is a subset of the dataset that is used to train the model. Testing Dataset: A dataset used to test the model is called the "testing dataset."

D. Creating the Modelthe Next Stage is to Design the Model

- Using various algorithms we constructed different models,
- Parameterizing the model
- Model compilation

E. Creating the Model

The next stage is to design the model.

- Using various algorithms we constructed different models,
- Parameterizing the model
- Model compilation
- Saving model for future prediction

F. Model Training and Testing

We utilized training data to develop the model, and we will employ a range of algorithms to iteratively validate the findings to ensure their precision.

G. Model Implementation

We have employed training data to develop the model, and we shall utilize a diverse array of techniques to evaluate the accuracy of the model's predictions. In the implementation of the model, we consider user inputs such as age, gender, body temperature, height, weight, duration of exercise, and heart rate, to predict the calories expended by an individual, with the resultant output being displayed on the screen.

Algorithm

These are the developmental steps:

STEP 1: Data preprocessing. The data undergoes preprocessing by converting categorical values of gender into numerical representations.

STEP 2: Data analysis must be conducted by plotting a correlation graph between the input parameters and calories.

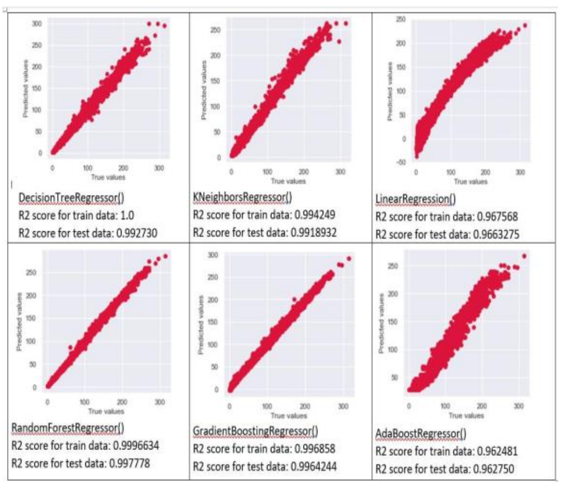
STEP 3: Training the model using the training data.

STEP 4: Evaluating the model with the assistance of testing data.

STEP 5: Preserve the model for prospective utilization.

H. Analysis of Various Regression Algorithms

The subsequent image illustrates the analysis of various regression algorithms.



[Fig. 1: Performance Analysis of Different Algorithms]

IV. METHODOLOGY

We noted minimal variance in the accuracy of training and testing data for random forests. Consequently, to forecast calories expended, we ultimately implemented the random forest algorithm within our application. Given that we already possess the user's height and weight, we classified the individual based on their BMI. Body Mass Index, or BMI, computes an individual's body fat percentage in relation to their height and weight. It represents a straightforward, expedient, and cost-effective methodology to screen for potential weight-related health issues, such as obesity [18].

Formula to determine BMI:

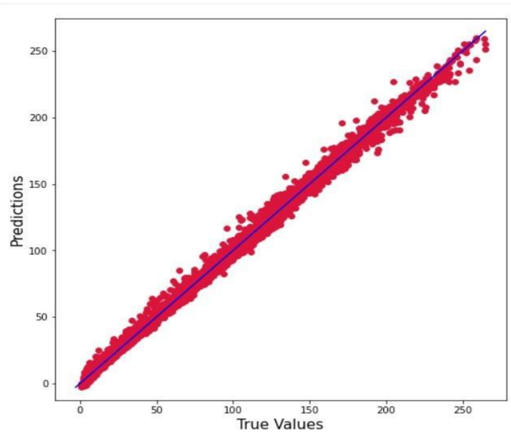
$$\text{BMI} = \frac{\text{weight (kg)}}{\text{height (m}^2\text{)}}$$

[Fig.2: BMI Formula]

Underweight	15–19.9
Normal weight	20–24.9
Overweight	25–29.9
Preobesity	
Class I obesity	30–34.9
Class II obesity	35–39.9
Class III obesity	≥40

Abbreviation: BMI, body mass index

[Fig.3: BMI Categories]



[Fig.4: Graph Between Results and Actual Values]

V. CONCLUSION

The objective of a calories burnt prediction system is to estimate the number of calories an individual will expend and to recommend a target for an individual to burn a specified quantity of calories to achieve weight loss or maintain health. This implementation endeavours to provide the most precise estimation of calories expended based on the user's inputs. It anticipates the individual's weight category according to their BMI [11], including classifications such as normal weight, underweight, obesity, and others. This also assists individuals in sustaining a nutritious diet and exercise regimen in order to uphold optimal health. Numerous previous studies encountered challenges with their findings, while others enhanced accuracy through the utilization of a more extensive dataset.

Retrieval Number: 100.1/ijsce.L931911111222
DOI: 10.35940/ijsce.L9319.14051124
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

By collecting data from a variety of disparate sources, the model's accuracy can be maximized.

VI. ACKNOWLEDGEMENTS

"This research is underpinned by the Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, India.."

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been sponsored or funded by any organization or agency. The independence of this research is a crucial factor in affirming its impartiality, as it has been conducted without any external sway.
- **Ethical Approval and Consent to Participate:** The data provided in this article is exempt from the requirement for ethical approval or participant consent.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Authors Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Liang, K.Y. and Zeger, S.L. (1993). Regression Analysis for Correlated Data. Annual Review of Public Health, 14(1), pp.43–68. https://doi.org/10.1146/annurev.pu.14.050193.000355
2. Tianqi Chen, Tong He "xgboost: eXtreme Gradient Boosting" Package Version: 0.6-4 January 4, 2017.
3. Najmeddine Dhieb1, Hakim Ghazzai, Hichem Besbes, and Yehia Massoud "Extreme Gradient Boosting Machine Learning Algorithm For Safe Auto Insurance Operations" 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES). Doi: https://doi.org/10.1109/ICVES.2019.8906396
4. Lin, W., Wu, Z., Lin, L., Wen, A. and Li, J. (2017). An Ensemble Random Forest Algorithm for Insurance Big Data Analysis. IEEE Access, 5, pp.16568–16575. https://doi.org/10.1109/ACCESS.2017.2738069
5. Stepanov, N., Alekseeva, D., Ometov, A. and Lohan, E.S. (2020). Applying Machine Learning to LTE Traffic Prediction: Comparison of Bagging, Random Forest, and SVM. 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). https://doi.org/10.1109/ICUMT51630.2020.9222418
6. Joshi, N., Singh, G., Kumar, S., Jain, R. and Nagrath, P. (2020). Airline Prices Analysis and Prediction Using Decision Tree Regressor. Data Science and Analytics, pp.170–186. https://doi.org/10.1007/978-981-15-5827-6_15
7. Tao Ban, Ruibin Zhang, Shaoning Pang, Abdolhossein Sarrafzadeh,and Daisuke Inoue "Referential kNN Regression for Financial Time Series Forecasting" National Institute of Information and Communications Technology 4-2-1 Nukui- Kitamachi, Tokyo, 184-8795, Japan. Doi: https://doi.org/10.1007/978-3-642-42054-2_75
8. A. Sen, M. Srivastava, Regression Analysis — Theory, Methods, and Applications, Springer-Verlag, Berlin, 2011 (4th printing). Doi: https://doi.org/10.1007/978-1-4612-4470-7
9. Rahmani, A.M. and Mirmahaleh, S.Y.H. (2021). Coronavirus disease (COVID-19) prevention and treatment methods and effective parameters: A systematic literature review. Sustainable Cities and Society, 64, p.102568. https://doi.org/10.1016/j.scs.2020.102568
10. Sagi, O. and Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. Information Sciences. https://doi.org/10.1016/j.ins.2021.05.055
11. Romero-Corral, A., Somers, V.K., Sierra-Johnson, J., Thomas, R.J., Collazo-Clavell, M.L., Korinek, J., Allison, T.G., Batsis, J.A., Sert-Kuniyoshi, F.H. and Lopez-Jimenez, F. (2008). Accuracy of body mass index in diagnosing obesity in the adult general population. International Journal of Obesity, [online] 32(6), pp.959–966. https://doi.org/10.1038/ijo.2008.11
12. Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G. (2020). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review. https://doi.org/10.1007/s10462-020-09896-5
13. Manikandan, S. (2010). Data transformation. Journal of Pharmacology and Pharmacotherapeutics, 1(2), p.126. https://doi.org/10.4103/0976-500X.72373
14. Eckerson, Wayne, W (2007). "Predictive Analytics. Extending the Value of Your Data Warehousing Investment" ISBN: 978-989-8533-58-6
15. Jiang, T., Gradus, J.L. and Rosellini, A.J. (2020). Supervised Machine Learning: A Brief Primer. Behaviour Therapy, 51(5), pp.675–687. https://doi.org/10.1016/j.beth.2020.05.002
16. DASH, M. and LIU, H. (1997). Feature selection for classification. Intelligent Data Analysis, [online] 1(1-4), pp.131–156. https://doi.org/10.3233/IDA-1997-1302 https://doi.org/10.1016/S1088-467X(97)00008-5
17. Srikanth, P., Anusha, C. and Devarapalli, D., 2015. A computational intelligence technique for effective medical diagnosis using decision tree algorithm. i-Manager's Journal on Computer Science, 3(1), p.21. https://doi.org/10.26634/jcom.3.1.3438
18. Nuttall, Frank Q. "Body Mass Index: Obesity, BMI, and Health: A Critical Review." Nutrition today vol. 50,3 (2015): 117-128. https://doi.org/10.1097/NT.0000000000000092
19. Sharma, P., & Site, S. (2022). A Comprehensive Study on Different Machine Learning Techniques to Predict Heart Disease. In Indian Journal of Artificial Intelligence and Neural Networking (Vol. 2, Issue 3, pp. 1–7). https://doi.org/10.54105/ijainn.c1046.042322
20. Yadav, R. K., Mishra, A. P., & Singh, A. (2021). Experimental Analysis of Covid 19 Spread Predictor using Linear Regression Algorithm. In International Journal of Innovative Technology and Exploring Engineering (Vol. 10, Issue 8, pp. 12–18). https://doi.org/10.35940/ijitee.h9076.0610821
21. Jain, N., & Kumar, R. (2022). A Review on Machine Learning & It's Algorithms. In International Journal of Soft Computing and Engineering (Vol. 12, Issue 5, pp. 1–5). https://doi.org/10.35940/ijsce.e3583.1112522
22. Sharma, Dr. N., & Iqbal, S. I. M. (2023). Applying Decision Tree Algorithm Classification and Regression Tree (CART) Algorithm to Gini Techniques Binary Splits. In International Journal of Engineering and Advanced Technology (Vol. 12, Issue 5, pp. 77–81). https://doi.org/10.35940/ijeat.e4195.0612523
23. Bhardwaj, S., Bhargava, Prof. N., & Bhargava, Dr. R. (2023). Genetic Algorithms: A Solution to Fiber Reinforced Composite Drilling Challenges. In International Journal of Emerging Science and Engineering (Vol. 11, Issue 6, pp. 1–5). https://doi.org/10.35940/ijese.f2548.0511623

AUTHORS PROFILE

Dr. Devarapalli Dharamaiah, is a Distinguished Professor in the Department of Computer Science and Engineering at Koneru Lakshmaiah University (KLU). With a wealth of knowledge and experience in the field, he specializes in areas such as software engineering, artificial intelligence, and data analytics. He is dedicated to fostering an innovative learning environment, mentoring students, and guiding research projects that contribute to advancements in technology. His commitment to academic excellence is reflected in his extensive publication record in reputed journals and conferences. Passionate about bridging the gap between theory and practice. He actively engages with industry partners to ensure that his students are well-prepared for the challenges of the tech world. Through his leadership and expertise, he continues to inspire future generations of engineers and researchers.

Ms. Yannam Satya Amrutha, holds a Bachelor's degree in Computer Science and Engineering from Shri Vishnu Engineering College for Women, where she excelled academically and gained hands-on experience through various technical projects.

She is deeply passionate about machine learning, deep learning, and data science, continually seeking to uncover the expansive potential within data-driven fields. Her learning journey is enriched by certifications, including Python for Everybody from Coursera, RDBMS, and Hadoop Fundamentals from IBM, reflecting her commitment to expand her expertise. Satya's dedication to continuous learning and innovation makes her a valuable asset in the tech industry.

Ms. Bala Satya Sri Pasupuleti, holds a Bachelor's degree in Computer Science and Engineering from Shri Vishnu Engineering College for Women in Bhimavaram, Andhra Pradesh, India. With a passion for technology, her skills span machine learning, deep learning, data science, and artificial intelligence. Bala is committed to academic excellence and is excited to publish her research, marking an important milestone in her journey. She actively explores real-world applications of her knowledge, seeking to contribute fresh insights to the field and positioning herself as a promising talent poised to make a lasting impact in technology.

Ms. Srujana Maddula, is a dedicated data scientist currently working at Target. She holds a degree in Computer Science from Shri Vishnu Engineering College for Women. With a strong foundation in machine learning and Android development, Srujana has contributed to many innovative projects. Her major work at Target was developing clustering components for an AI markdown recommendation system. Her passion for technology and data-driven solutions drives her to continuously explore and implement advanced analytical and development techniques in her work. In her free time, she enjoys writing about technology on various popular blogs, and sharing her insights and expertise with a wider audience.

Ms. Hema Sri Puppala, completed her Bachelor of Technology degree in Computer Science and Engineering from Shri Vishnu Engineering College for Women, Andhra Pradesh. Hema is passionate about exploring advancements in technology and programming. With a strong interest in software development, she actively pursues opportunities to expand her knowledge and gain hands-on experience. Hema values collaboration, engaging actively with peers and faculty to foster learning. As she advances in her studies, Hema is eager to make a positive impact in the tech industry and build a successful career in technology.
