

# Burned Calories Prediction: Burned Calories Prediction using Machine Learning

Abdallah Hamdy Shaaban<sup>1</sup>, Mohamed Ahmed Mohamed<sup>2</sup>,  
Daniel Michel Iskandar<sup>3</sup>, Abdulrahman Abdalmoniem Ahmed<sup>4</sup>, Ahmed khaled Maher Hassan<sup>5</sup>

*Faculty of Computer Science*

*Misr International University, Cairo, Egypt*

abdallah2110678<sup>1</sup>, <sup>2</sup>,

mohamed2110111<sup>3</sup>, daniel2101026<sup>4</sup>, Abdulrahman2111656<sup>5</sup>, ahmed2003973<sup>6</sup>{@miuegypt.edu.eg}

**Abstract**—Regular physical activity is essential to maintaining health and fitness. Estimates of personal calorie expenditure are based on training formulas. Today, people are looking for quick solutions to all their problems. I want to reduce the amount of exercise and get more results in order to check the degree of improvement and calorie consumption of the human body after exercise. In this paper propose predict by machine learning algorithms Models which takes some attributes as input and gives approximate calories burnt value which will motivate people to do more exercise. Regression Model training and testing using Linear Regression, Decision Tree Regressor, Random Forest Regressor, AdaBoost Regressor, Gradient Boosting Regressor and XGBoost Regressor were done to determine the best model for the study. XGBoost Regressor achieved a lower Mean Squared Error (4.81), Root Mean Squared Error (2.19), and Mean Absolute Error (1.51), coupled with a higher R-squared value (0.999). Visualizing and exploring the relationships between variables in your data is very important because it can affect the performance of an algorithm in predicting the value of a target variable. The Algorithm model was able to predict the calories burned with a high accuracy rate.

**Keywords:** burnt calorie; Machine Learning; regression ; Ada boost regression; Linear Regression

## I. INTRODUCTION

Calories burned depend on internal and external factors, is subjective, and varies from person to person depending on height, weight, and fitness level. Calorie generally refers to the loss of weight or food, but it is the amount of heat energy. From a human perspective, the number of calories is the amount of energy required to perform a task. Different foods have different calorie values.

- 1) **Age (years)** – As you get older, your body contains fewer calories. This is often due to decreased physical activity, changes in metabolism, or age-related loss of bone and muscle mass.
- 2) **Height** – Taller people have more muscle mass, which is related to metabolic rate. This means that a person needs more calories to function. Therefore, taller people burn more calories than shorter people.
- 3) **Weight** – If you take in more calories than you burn, you will gain weight. This means that the more you weigh, the more calories you need to burn to balance your calorie intake.

- 4) **Gender** – The human body requires different calories depending on gender. Women need fewer calories than men because men have more muscle mass.
- 5) **Heart rate** – During exercise, your heart rate increases, indicates the effort you put into performing the activity, and determines the calories you burn. A higher heart rate also helps your body burn fat, which is the building block for calories.
- 6) **Body temperature** – When your body temperature increases, more energy is used to cool it down, burning calories.
- 7) **Duration** – The longer you exercise, the higher your body temperature will be and the more calories you will burn. Therefore, the duration of exercise directly affects the amount of calories burned in the body.



Although there are several studies using machine learning techniques to predict calories burnt, there are still significant gaps in the literature. Most of the existing research has focused on predicting calorie consumption during specific types of physical activity or in specific populations. Generalized models that can accurately predict calorie burn across a variety of physical activities and individuals are needed. The main objectives of these studies are: Collect data on physical activity and calories burnt from a variety of sources, including fitness trackers and wearable devices. Data must be preprocessed and cleaned to ensure accuracy and consistency.

When the human body engages in extensive activity and exercise, body temperature and heart rate begin to increase, resulting in the production of thermal energy within the body,

that leads to burning calories. To demonstrate the same, we took input parameters like age, gender, height, weight, etc. and ran various regression algorithms like linear regression, XG boost regression, AdaBoost regression, SVR, decision tree regression, random forest regression, etc. on the data.

Using previously collected data from more than 6,400 subjects ranging from 8 days old to 95 years old, the researchers found that, taking into account body size and amount of fat and muscle, human metabolism generally changes throughout life. They found that it goes through four distinct stages. The metabolism of newborns is similar to that of adults. Then, around the first month of life, the metabolism begins to increase rapidly, and by the age of 9 to 15 months his metabolism is more than 50 percent higher than that of an adult. This is equivalent to an adult consuming about 4,000 calories per day. (The U.S. Department of Health and Human Services estimates that, on average, an adult woman needs 1,600 to 2,400 calories per day and an adult man needs 2,000 to 3,000 calories per day. )



The research paper on Calories Burnt Prediction using Machine Learning is organized into four distinct sections. Section 1 provides a comprehensive review of related work, background information, and a literature review, contextualizing the study within existing methodologies and advancements in predicting calorie expenditure. Section 2 outlines the proposed algorithms alongside a detailed description of the datasets used, explaining the methodologies and dataset characteristics pivotal for predicting calories burnt. Section 3 encompasses the results, utilizing a combination of images, tables, and textual analysis to present empirical findings derived from applying the algorithms to the datasets. This section aims to showcase the performance metrics and comparisons between predicted and actual calorie expenditure. Finally, Section 4 engages in a comparative discussion, analyzing why specific algorithms outperform others based on metrics like accuracy, robustness, and practical applicability, contributing significant insights to the field of calorie prediction using machine learning techniques.

## II. RELATED WORK

The field of calories burned prediction is not unexplored; a lot of people investigated the field and came up with satisfactory results. We will mention some of the papers we read and analyzed to assist us in our research, and all the papers mentioned will be referenced in the references section.

Marte Nipas1 et al.[1] The paper "Burned Calories Prediction using Supervised Machine Learning: Regression Algorithm" presents a study on predicting calories burned using supervised machine learning regression algorithms. The study uses data from the Kaggle website, consisting of 15000 observations and nine variables. The authors analyze the dataset, extract features, and train three regression models: linear regression, ridge regression, and random forest regression. They use K-fold cross-validation to test the models and determine the best model for predicting calories burned. The random forest regression model achieves the highest accuracy of 95.77%. The paper emphasizes the importance of data analysis, model training, and visualization of variable relationships. It also suggests potential areas for further research and acknowledges the Technological Institute of the Philippines High-Performance Computing Laboratory for providing computing facilities.

Punita Panwar et al.[2]This study, "Study on Calorie Expenditure Prediction Using Machine Learning," focuses on predicting calorie expenditure during exercise using machine learning techniques. The team collected data from a Kaggle repository using two CSV files containing 15,000 instances of seven attributes including gender, age, height, weight, body temperature during exercise, heart rate, and duration of exercise. The analysis included the application of various machine learning models, and he identified XGBoost regression as the best algorithm for predicting calorie consumption, achieving a mean absolute error (MAE) of 1. The study aimed to motivate individuals to exercise more and live healthier lives by providing accurate predictions of calorie consumption. The study also highlighted the importance of understanding the factors that influence calorie expenditure and the importance of accurate predictions in promoting fitness and health.

This paper presents a machine learning-based portable system for determining eating habits of a human being using a six-point calibrated wearable MEMS tri-axial accelerometer. The system consists of a wrist-worn MEMS accelerometer that calculates calories burned per step, which is directly sent over to the user's smartphone and a cloud-based machine learning algorithm that does the prediction of health habit (i.e. healthy, unhealthy or undernutrition) based on the data obtained from the wrist-worn device. The proposed system not only decreases the risk of obesity in a person but also introduces a low-cost alternative device with reduced power consumption and minimal covering size that can improve

people's life.

Thomas Reichherzer et al.[3] This article describes security and privacy vulnerabilities associated with wearable fitness devices and how supervised machine learning techniques can be used to collect and analyze data from these devices to track people and make predictions.to introduce.The study shows that these methods can violate user privacy, and the paper provides recommendations to improve the security of wearable devices based on the findings.

This research article focuses on machine learning-based feature selection intent and inference of calories burned during exercise. The team uses a training dataset containing eight independent variables and one dependent variable, preprocesses the dataset using feature scaling and missing values, performs exploratory feature testing, and analyzes the raw dataset. We address this issue by fitting all regressors and subjecting the raw dataset to the following tests: Analysis of performance using feature selection axioms and metrics such as MAE, MSE, EVS, RScore, and runtime. The results have implications for computer science and mathematics education.

The research a machine learning-based approach to estimate energy expenditure (EE) during aerobic exercise, taking into account individual characteristics and excess post-exercise oxygen consumption (EPOC).This study included his 33 subjects and proposed a model that achieved high accuracy in estimating EE.The model takes into account characteristics such as training duration, intensity, gender, age, height, weight, and heart rate.The results showed that machine learning-based approaches, especially linear regression and Gaussian processes, provide good approximations to the real EE.This study also highlighted the importance of considering his EPOC in estimating EE and its significant contribution to total EE. Potential applications of this model include integration into wearable devices such as smartwatches for sports training,weight management, and health monitoring.

Niharikareddy Meenigea et al.[4]This research on predicting calorie consumption using machine learning models, specifically the XGBoost regressor algorithm. The study takes into account physiological and environmental factors such as exercise duration, heart rate, body temperature, height, weight, age, and gender. The dataset used in the study contains more than 15,000 data points, and the average absolute error of the model is 2.7. The study also includes data analysis and visualization techniques to evaluate model performance. The XGBoost regressor algorithm has been proven to be an effective and efficient method for predicting calorie consumption. The findings can also be applied to other health-related predictions, such as tracking sleep patterns and heart rate.

Gurrappagaru Sanjana Reddy et al.[5] You need to know an accurate estimate of how many calories you burn during

physical activity. This plays an important role in improving an individual's health status. Machine learning, in particular he explores the use of the XGBoost regression algorithm to provide accurate calorie expenditure predictions. In this study, we used various physical activity and biometric data to build a robust predictive model using multiple algorithms such as XGBoost regressor, linear regression, decision tree regressor, and random forest regressor. The XGBoost regression algorithm consistently outperforms other algorithms and provides the most accurate and reliable calorie expenditure estimates.

Sona P Vinoy et al.[6] prediction of the amount of calories burned during exercise for individuals aiming to maintain, lose, or manage weight.compare the performance of XGBoost Regressor and Linear Regression algorithms in predicting the calories burned during exercise.The dataset is consists of 7 features, one target variable, and 15000 instances then divided into training and test sets. XGBoost Regressor and Linear Regression are used as machine learning models for comparison.XGB Regressor with The mean absolute error getting is 2.71 rather that Linear Regression with The mean absolute error getting is 8.38.

Rachit Kumar Singh et al.[7] predict the number of calories burnt during exercise using machine learning techniques.create a model for prediction calories burnt during exercise, considering factors such as duration of exercise, average heart rate, height, weight, and gender of the individual. use of Xgboost Regressor, Linear Regression, Logistic Regression, and Lasso Regression models. the mean absolute error for Lasso Regression is 2.7159,as well as evaluate the models based on test data and comparing the data with original values.

Rahul katarya et al.[8] The paper describes the use of machine learning techniques to predict calorie expenditure during exercise, with particular emphasis on the Xgboost regression model. Describes the use of various regression techniques such as linear regression, logistic regression, and Lasso regression to build predictive systems. The paper also describes the relationship between exercise, heart rate, body temperature, and energy expenditure. Additionally, it provides methodologies for data analysis, model training, and evaluation. The conclusion highlights the effectiveness of his Xgboost regressor for this prediction task. The reference section contains a variety of resources on machine learning and data science.

This study, "Calorie consumption prediction using machine learning," aims to predict an individual's calorie consumption using the XG-Boost regression model. The study involves collecting and preprocessing a dataset from her Kaggle that includes attributes such as height, weight, gender, age, exercise time, heart rate, and body temperature. The XG-Boost regression model was fed over 15,000 data instances

and reported a mean absolute error of 2. This study highlights the importance of accurately predicting calorie consumption to support weight management and public health efforts. Additionally, the relationship between various factors such as exercise duration, heart rate, and body temperature and calorie expenditure has been investigated. The results of this study have potential implications for individuals seeking to manage calorie intake and burn to improve health and fitness.

### III. PROPOSED METHODOLOGY

Numerous algorithms were used, and a research was done on each algorithm before training the model using them on the datasets. The following diagram represents the steps the datasets went through to get the results.

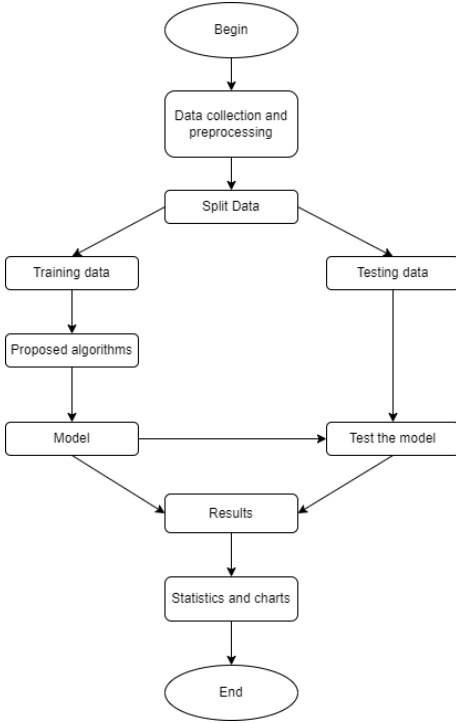


Fig. 1. Heart disease prediction process

#### A. Datasets Descriptions

The dataset for predicting calories burned comprises 7 key features and encompasses 13,000 records, partitioned into 90

Gender: Denotes the gender of the individual. Age: Represents the age of the individual. Height: Specifies the height of the individual. Weight: Reflects the weight of the individual. Duration: Indicates the duration of the exercise session. Heart Rate: Represents the heart rate during exercise. Body Temperature: Specifies the body temperature during exercise. These attributes are essential in estimating the calorie expenditure accurately. In this context, 'Gender' differentiates between male and female subjects, 'Age' accounts for age-related factors impacting calorie burn, 'Height' and 'Weight' play roles in basal metabolic rate calculations, 'Duration' of

exercise sessions influences calorie burn, while 'Heart Rate' and 'Body Temperature' provide insights into physiological responses affecting energy expenditure.

TABLE I  
FEATURES OF DATASET 1

Feature	Type	Values
Gender	Classification	Male or Female
Age Category	Classification	Intervals of age(80 or older, 55-59, 65-69, etc.)
Height	Numerical	from 150 to 209
Weight	Numerical	from 50 to 109
Duration	Numerical	From 0 to 30
Heart <sub>Rate</sub>	Numerical	From 70 to 120
Body <sub>Temp</sub>	Numerical	From 37.8 to 40.9

#### B. Used Algorithms

The mentioned datasets were passed into 6 different Machine Learning algorithms which were XGB Regression, Ag Boost ,Ada Boost Regression, Random Forest, Decision Tree, and Linear Regression. For each of the algorithms there were statistics generated, these statistics were: Accuracy, Recall, Precision, and Specificity. Afterwards, the results were charted and compared. The results, charts, and the discussion of the results can be found later in the paper.[9]

#### XGBRegressor Algorithm:

Formal Parameters:

- *n\_estimators*: Number of boosting rounds (trees) to be built.
- *max\_depth*: Maximum depth of each tree.
- *learning\_rate*: Rate at which the model adapts through boosting.
- *subsample*: Fraction of samples used for each boosting round.
- *col\_sample\_bytree*: Fraction of features used for each boosting round.
- *objective*: Loss function to be minimized.

Algorithm:

#### XGBRegressor Algorithm:

Formal Parameters:

- *n\_estimators*: Number of boosting rounds (trees) to be built. It represents the total number of sequential trees constructed during the boosting process.
- *max\_depth*: Maximum depth of each tree. This parameter controls the depth of each tree in the boosting process, limiting the number of nodes.
- *learning\_rate*: Rate at which the model adapts through boosting. It scales the contribution of each tree added to the model.
- *subsample*: Fraction of samples used for each boosting round. It specifies the fraction of the training data to be used for each boosting iteration.
- *col\_sample\_bytree*: Fraction of features used for each boosting round. It denotes the fraction of features to be randomly selected for each tree.



- **objective:** Loss function to be minimized. It defines the loss function to be optimized during training, such as 'reg:squarederror' for regression tasks.

**Algorithm Explanation:**

- 1) **Initialization:** Initialize the XGBRegressor model with default parameters or user-defined values for the formal parameters.
- 2) **Boosting Rounds:** For each boosting round:
  - a) **Gradient Calculation:** Calculate the gradient of the loss function with respect to the predictions.
  - b) **Tree Construction:** Build a decision tree to minimize the calculated gradient, using parameters like *max\_depth*, *subsample*, and *col\_sample\_bytree* to control tree complexity and data subsets.
  - c) **Model Update:** Update the model by adding the predictions of the new tree to the ensemble with a scaled value (learning rate).
- 3) **Return:** Return the trained XGBRegressor model after completing the specified number of boosting rounds.

### Decision Tree:

The supervised learning type includes the decision tree algorithm. Both regression and classification issues may be handled using them. Each node in the tree corresponds to a class label, with attributes expressed on the tree's inner node. Any Boolean function with discrete characteristics may be described using the decision tree. The entropy varies when a node is employed in a decision tree, it breaks down the training dataset into smaller groupings. the increase in entropy is denoted by the information. [10].

**Definition:** Suppose  $S$  is a set of instances,  $A$  is an attribute,  $S_v$  is the subset of  $S$  with  $A = v$ , and  $\text{Values}(A)$  is the set of all possible values of  $A$ , then

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \left| \frac{S_v}{S} \right| \cdot \text{Entropy}(S_v) \quad (1)$$

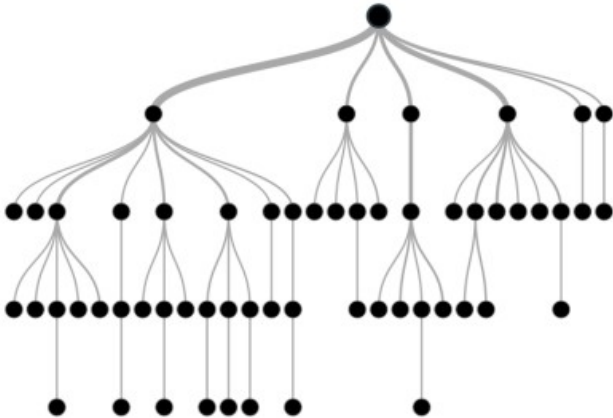


Fig. 2. Illustration of decision tree[11]

### AdaBoost Algorithm:

**Formal Parameters:**

- $T$ : Number of iterations (weak learners).
- $D_t$ : Distribution over the training samples at iteration  $t$ .
- $h_t(\mathbf{x})$ : Weak learner (classifier) at iteration  $t$ .
- $\alpha_t$ : Weight of the weak learner  $h_t(\mathbf{x})$  in the final prediction.
- $\epsilon_t$ : Weighted error of the weak learner  $h_t(\mathbf{x})$ .

**Algorithm Explanation:**

- 1) **Initialization:** Initialize the sample weights  $D_1(i) = 1/N$  for  $i = 1, 2, \dots, N$ , where  $N$  is the number of training samples.
- 2) **Iteration:** For  $t = 1$  to  $T$ :
  - a) **Train Weak Learner:** Train a weak learner  $h_t(\mathbf{x})$  using the distribution  $D_t$ . The weak learner focuses on samples with higher weights.
  - b) **Calculate Weighted Error:** Compute the weighted error  $\epsilon_t$  of  $h_t(\mathbf{x})$  based on misclassifications.
  - c) **Calculate Learner Weight:** Calculate  $\alpha_t = \frac{1}{2} \log \left( \frac{1 - \epsilon_t}{\epsilon_t} \right)$ , which represents the weight of  $h_t(\mathbf{x})$  in the final prediction.
  - d) **Update Weights:** Update sample weights  $D_{t+1}(i) = \frac{D_t(i) \cdot \exp(-\alpha_t y_i h_t(\mathbf{x}_i))}{Z_t}$ , where  $y_i$  is the true label of sample  $i$  and  $Z_t$  is a normalization factor.
- 3) **Output:** Construct the final strong learner by combining the weak learners weighted by  $\alpha_t$ .

### Linear Regression:

Linear regression is an algorithm that provides a linear relationship between an independent variable and a dependent variable to predict the outcome of future events. It is a statistical method used in data science and machine learning for predictive analysis. Thus, linear regression is a supervised learning algorithm that simulates a mathematical relationship between variables and makes predictions for continuous or numeric variables[12]

$$Y = m \cdot X + b \quad (2)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^n (y_i - (mx_i + b))^2 \quad (3)$$

### Random Forest:

Random Forest is one of the supervised machine learning algorithms that can be used for both classification and regression tasks but, it works better in classification tasks. This algorithm considers multiple decision trees before giving an output. This technique is founded on the notion that a greater number of trees would ultimately guide to the rectified selection. It employs a voting approach for classification and then determines the class, whereas it uses the mean of all the decision tree outputs for regression[13]. Random Forest

Algorithm is extremely efficient with large datasets with high dimensionality. [14]

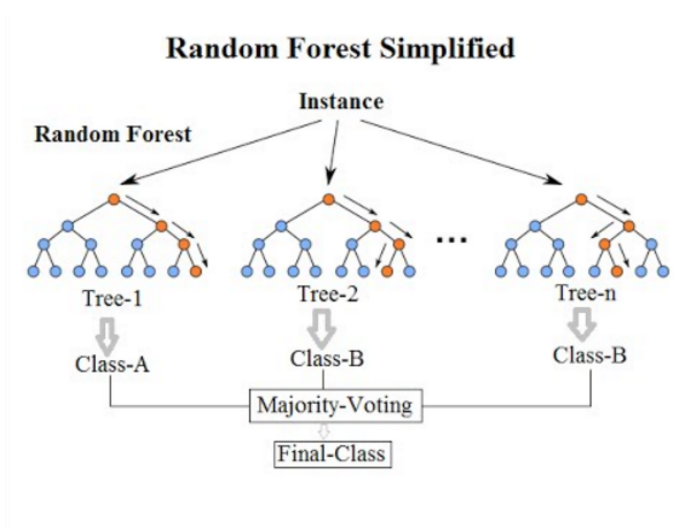


Fig. 3. Random forest demonstration [11]

### Gradient Boosting Regressor Algorithm:

Formal Parameters:

- *n\_estimators*: Number of boosting stages (trees) to be built.
- *learning\_rate*: Rate at which the boosting algorithm learns from mistakes.
- *max\_depth*: Maximum depth of the individual regression estimators.
- *subsample*: Fraction of samples used for fitting the individual base learners.
- *loss*: Loss function to be optimized during training. Common choices include 'ls' for least squares regression.

Algorithm Explanation:

- 1) **Initialize the Model:** Initialize the Gradient Boosting Regressor with specified or default parameter values.
- 2) **Iteration:** For each boosting stage (from 1 to *n\_estimators*):
  - a) **Fit a Tree:** Train a regression tree using the current residuals of the predictions.
  - b) **Update Predictions:** Update the model's predictions by adding the predictions of the new tree, scaled by the learning rate.
- 3) **Return:** Return the trained Gradient Boosting Regressor model after all boosting stages.

## IV. RESULTS AND ANALYSIS

### A. Overall Model Performance

#### B. Decision Tree Regressor

##### 1) Training Metrics:

#### C. Random Forest Regressor

##### 1) Training Metrics:

Model	Mean Squared Error	RMSE	R-squared	Mean Absolute Error
Linear Regression	505.99	22.49	0.87	17.89
Decision Tree Regressor	391.24	19.78	0.90	14.06
Random Forest Regressor	222.52	14.92	0.94	10.52
AdaBoost Regressor	384.57	19.61	0.90	16.41
Gradient Boosting Regressor	201.91	14.21	0.95	10.13
XGBoost Regressor	204.20	14.29	0.95	10.12

TABLE II

PERFORMANCE METRICS FOR DIFFERENT REGRESSION MODELS

Metric	Mean Squared Error	RMSE	R-squared	Mean Absolute Error
Training	0.95	0.97	0.99	0.06
Validation	391.24	19.78	0.90	14.06

TABLE III

DECISION TREE REGRESSOR METRICS

#### D. AdaBoost Regressor

##### 1) Training Metrics:

#### E. Gradient Boosting Regressor

##### 1) Training Metrics:

#### F. XGBoost Regressor

##### 1) Training Metrics:

#### G. Cross-Validation

##### 1) 2-fold Cross-Validation:

2) **10-fold Cross-Validation:** Based on the numbers, XGBoost Regressor appears to be the best algorithm for this specific predictive task as it achieved the lowest errors and highest R-squared value, indicating superior predictive accuracy and overall performance compared to the other models.

Using k-fold on the dataset had slight improvements with most algorithms. Logistic Regression had 0.009 of increased precision. Gradient Boosting suffered a very slight loss of 0.001 in terms of specificity. Random Forest, Decision Tree, and Naive Bayes all suffered losses in terms of accuracy. K-NN had a slightly decreased specificity compared to the 70/30 data split.

The following results are from the second dataset. Surprisingly, Gradient Boosting had the best accuracy in this dataset scoring a 0.962. Results were better for almost all the algorithms, but that could be attributed to the significantly smaller dataset. All the algorithms had a much higher specificity. Logistic Regression lost the first place for the first time in our testing, scoring a mediocre 0.843 in terms of accuracy.

TABLE X

STATISTICS OF ALGORITHMS WITH 10 K-FOLD

Model	Accuracy	Precision	Recall	Specificity
Logistic Regression	0.849	0.853	0.849	0.844
Gradient Boosting	0.965	0.965	0.965	0.965
k-NN	0.847	0.848	0.847	0.844
Random Forest	0.986	0.986	0.986	0.986
Decision Tree	0.951	0.952	0.951	0.952
Naïve Bayes	0.849	0.849	0.849	0.848

Metric	Mean Squared Error	RMSE	R-squared	Mean Absolute Error	Metric	Mean Squared Error	RMSE	R-squared	Mean Absolute Error
Training	31.98	5.65	0.99	3.99	Training	192.30	13.87	0.95	10.11
Validation	222.52	14.92	0.94	10.52	Validation	201.91	14.21	0.95	10.13

TABLE IV  
RANDOM FOREST REGRESSOR METRICS

TABLE VI  
GRADIENT BOOSTING REGRESSOR METRICS

Metric	Mean Squared Error	RMSE	R-squared	Mean Absolute Error	Metric	Mean Squared Error	RMSE	R-squared	Mean Absolute Error
Training	374.05	19.34	0.90	16.26	Training	121.11	11.01	0.97	7.89
Validation	384.57	19.61	0.90	16.41	Validation	204.20	14.29	0.95	10.12

TABLE V  
ADABOOST REGRESSOR METRICS

TABLE VII  
XGBOOST REGRESSOR METRICS

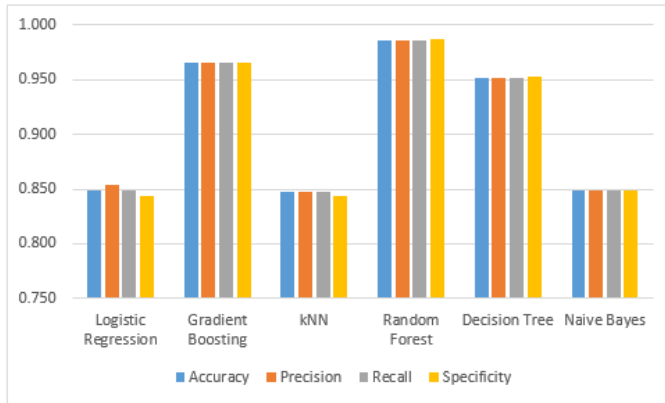


Fig. 4. Third dataset performance chart with 10 k-fold

Using k-fold, this testing had the highest accuracy. Random Forest scored an accuracy of 0.986. Gradient Boosting came second with an accuracy of 0.965. As seen in the previous testing specificity was much higher in all the algorithms.

## V. CONCLUSION

An early prediction of calorie expenditure is crucial for individuals seeking to maintain a healthy lifestyle, and machine learning has proven to be a powerful tool in accurately estimating calorie burn. Through extensive research and collaboration with experts in this field, the predictive capabilities of machine learning algorithms, such as XGBoost Regressor, have proven to be significantly accurate. They outperform other algorithms that XGBoost Regressor achieved a lower Mean Squared Error(4.81), Root Mean Squared Error (2.19), and Mean Absolute Error (1.51), coupled with a higher R-squared value (0.999). Machine Learning can be applied to many fields, not just calorie burnt, it can be used to predict anything from stock prices to results of sports matches which makes it a very useful tool for humanity. And this tool will only keep improving and producing better results.

## VI. ACKNOWLEDGMENT

First and foremost, we would like to show our gratitude for the staff working in Misr International University

and the faculty of computer science for working hard to make this university a success. We are very thankful for Prof. Mohamed Shebl El Komy University President, Prof. Ayman Bahaa dean faculty of Computer Science, Prof. Abdelnasser Zaied Vice Dean of Student Affairs and Professor of Computer Engineering, and Dr. Ayman Nabil Associate Professor in Computer Science for giving us the opportunity to learn in this virtuous university and running it flawlessly. And finally, we would like to give special thanks to Dr. Daaa AbdelMoneim associate professor in information systems and Eng. Mostafa Radwan teaching assistant for their continued guidance and support for our work.

## REFERENCES

- [1] J. N. M. M. A. F. M. J. A. B. S. J. S. G. Marte Nipas, Aimee G. Acoba, "Burned calories prediction using supervised machine learning: Regression algorithm," 2022.
- [2] R. s. a. R. S. Punita Panwar, Kanika Bhutani, "A study on calories burnt prediction using machine learning," vol. 54, 2023.
- [3] N. E. N. R. Thomas Reichherzer, Mikayla Timm, "Using machine learning techniques to track individuals their fitness activities," 2017.
- [4] N. Meenigea, "Calorie burn prediction: A machine learning approach using physiological and environmental factors," vol. 1, pp. 587–595, 2014.
- [5] S. V. S. Gurrappagaru Sanjana Reddy, "Accurate calorie burn prediction with machine learning: Xgboost in focus," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 5, pp. 806–812, 2023.
- [6] B. J. Sona P Vinoy, "Calorie burn prediction analysis using xgboost regressor and linear regression algorithms," vol. 4, pp. 187–191, 2022.
- [7] V. G. Rachit Kumar Singh, "Calories burnt prediction using machine learning," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 11, pp. 577–588, 2022.
- [8] R. Katarya and S. K. Meena, "Machine learning techniques for heart disease prediction: a comparative study and analysis," *Health and Technology*, vol. 11, no. 1, pp. 87–97, 2021.
- [9] A. Singh and R. Kumar, "Heart disease prediction

Model	Average MSE
XGBoost Regressor	Lowest among all models

TABLE VIII  
2-FOLD CROSS-VALIDATION RESULTS

TABLE IX  
10-FOLD CROSS-VALIDATION RESULTS

- using machine learning algorithms,” in *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, 2020, pp. 452–457.
- [10] M. Shouman, T. Turner, and R. Stocker, “Using decision tree for diagnosing heart disease patients,” in *Proceedings of the Ninth Australasian Data Mining Conference-Volume 121*, 2011, pp. 23–30.
  - [11] V. Ramalingam, A. Dandapath, and M. K. Raja, “Heart disease prediction using machine learning techniques: a survey,” *International Journal of Engineering Technology*, vol. 7, no. 2.8, pp. 684–687, 2018.
  - [12] B. Deekshatulu, P. Chandra *et al.*, “Classification of heart disease using k-nearest neighbor and genetic algorithm,” *Procedia technology*, vol. 10, pp. 85–94, 2013.
  - [13] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Prediction of heart disease using random forest and feature subset selection,” in *Innovations in bio-inspired computing and applications*. Springer, 2016, pp. 187–196.
  - [14] S. Asadi, S. Roshan, and M. W. Kattan, “Random forest swarm optimization-based for heart diseases diagnosis,” *Journal of Biomedical Informatics*, vol. 115, p. 103690, 2021.