CALORIE BURNT PREDICTION USING MACHINE LEARNING

K.V.Narasimha Reddy
Asst.Professor
Computer Science & Engineering
Narsaraopeta Engineering College
(Autonomous)
Narasaraopeta,India
narasimhareddynec03@gmail.com

Dharma Teja Grandhe
Student
Computer Science & Engineering
Narsaraopeta Engineering College
(Autonomous)
Narasaraopeta,India
grandhi.dharmateja@gmail.com

Aravind Gorantla
Student
Computer Science & Engineering
Narsaraopeta Engineering College
(Autonomous)
Narasaraopeta,India
aravindchow1407@gmail.com

Raju Konna
Student
Computer Science & Engineering
Narasaraopeta Engineering College
(Autonomous)
Narasaraopeta,India
rajukonna2894@gmail.com

Satish Kumar Naraboyina
Student
Computer Science & Engineering
Narasaraopeta Engineering College
(Autonomous)
Narasaraopeta,India
satishnaraboyina@gmail.com

Abstract--Consistent participation in physical activities is vital for preserving overall health and fitness. Our research is centered on employing regression models within machine learning to forecast the calories expended during these activities with precision. Our goal is to generate accurate results for calorie expenditure through the utilization of these models. The dataset includes variables like heart rate, duration of activity, age, gender, weight and type of exercise which helps in developing a better and accurate prediction model. Before feeding the data to the model primary steps to follow. Data preprocessing, cleaning, and analysis of the data. Also, splitting into train and test data to train the models and validate them. By using different regression models getting the best accurate prediction for the calories burned.

Keywords -- Regression models, Machine learning, Calories, Prediction.

I. INTRODUCTION:

A measure of energy obtained from the consumption of edibles and drinks, as well as the energy expended during physical activities, is known as a calorie [1]. Food items vary in their caloric content, influencing the energy they provide. Therefore, managing calorie intake is crucial to mitigate health issues like obesity, diabetes, and related conditions [2].In contemporary living, maintaining regular physical activity is essential for achieving fitness goals, promoting overall health, and managing weight. Activities such as running, walking, cycling, swimming, exercising, and even routine tasks contribute to calorie expenditure [3]. However, the actual number of calories expended relies on personal aspects like age, metabolic rate, gender, height and the level of exertion in the exercise. While accurately quantifying calorie burn can be challenging, various methods such as activity tracker apps and heart rate monitoring offer estimations, albeit with differing levels of accuracy [4]. Additionally, the Metabolic Equivalent of Task (MET) system [5], developed by researchers and endorsed by the medical community, provides a standardized approach to estimating calorie expenditure based on activity intensity.

To address the challenge of predicting calorie burn, this study aims to utilize machine learning algorithms. Specifically, Linear Regression, Ridge Regression, XGBoost Regression, and Random Forest Regression will be considered. The objective is to determine which algorithm can provide the most precise expectations of calories burned relies on attributes such as heartrate, age, activity term, weight, height, and temperature.

The coming about could be integrated into existing technologies to enhance the precision of calorie burn estimates following physical activities.

II. LITERATURE SURVEY

Machine learning algorithms have seen widespread adoption in recent years for predicting calorie burn during physical activity. These studies typically gather data from fitness trackers, mobile apps, and wearable devices, including physical activity data, heart rate, age, and gender. Here's an overview of key research in this field:

Sathiya T et al. [6] proposed a method to predict users' calorie intake by applying a CNN model to classify food items from input images. They achieved a 91.65% accuracy in predicting calorie intake using deep learning and image processing techniques.

Sona P Vinoy et al. [7] focused on predicting calorie burn during workouts. They utilized machine learning calculations like XGBoost regressor and straight relapse models, utilizing traits such as age, tallness, weight, term, heart rate, body temperature, and calorie. While they didn't report model accuracy, their mean absolute error was approximately 2.71 for XGBoost and 8.31 for linear regression, based on a dataset of 15,000 CSV entries.

Suvarna Shreyas Ratnakar et al. [8] also aimed to forecast the calories expended from physical exercises. They used the XGBoost algorithm on a dataset of 15,000 entries, achieving a mean absolute error of 2.7. However, model accuracy wasn't mentioned.

Rachit Kumar Singh et al. [9] presented a method for predicting calorie burn using logistic regression, linear regression, and lasso regression models.

Unfortunately, they didn't provide details on error metrics, dataset size, or model accuracy.

Marte Nipas et al. [10] utilized a supervised learning algorithm, Random Forest, to predict burned calories with an impressive model accuracy of 95.77%. They also employed an iterative method for output determination, showcasing superior performance compared to other recent studies.

Gunasheela B L et al. [11] outlined their approach for predicting calorie intake from input images using digital image processing techniques, including image acquisition, RGB conversion, feature extraction, and image enhancement. They segmented input images and applied various techniques to predict calorie intake.

KR Westerterp et al. [12] discussed determining energy expenditure based on body size, composition, food intake, and physical activity, employing statistical techniques. However, specific details about their methods and results were not provided.

In summary, these investigations underscore the promise of machine learning techniques in accurately predicting energy expenditure during physical activity. Nonetheless, there remains a need for models capable of accurately predicting energy expenditure across diverse physical activities and individuals.

Additionally, as machine learning gets better, it can help create personalised advice to improve energy usage and overall health. To make this happen, future research needs to focus on making models that can handle the differences between individuals and activities they do.

III. METHODOLOGY

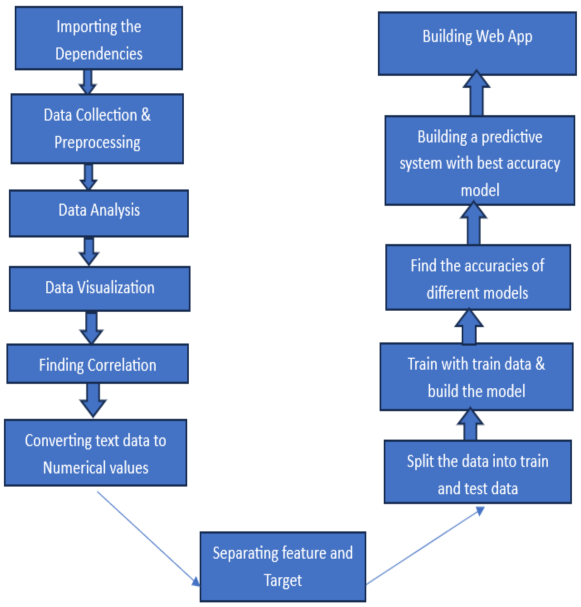


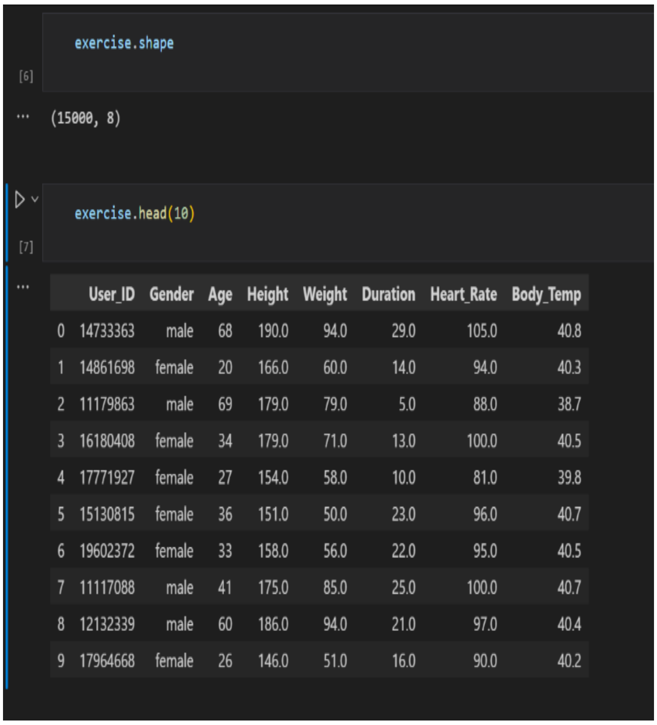
Fig1: Work flow of Calorie Burnt Prediction

Research Objective: The primary objective of this research is to employ machine learning methodologies for predicting how many calories you burn during physical activities. The basic workflow diagram is shown in above figure 1.

A. *Data Description:*

Getting the data is a crucial step in any machine learning project, as the quality of the data affects how well the model works. In this study, we got our data from Kaggle, a popular website where data scientists share datasets. After collecting the data, which had over 15,000 records and 7 different pieces of information, we put it onto Google Collab, an online platform for analysing data and doing machine learning.

In figure2 the 7 input attributes gender, age, height, weight, duration, heartrate, body_temp and the size of the dataset is displayed.



```
[6] exercise.shape
(15000, 8)
[7] exercise.head(10)
```

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp
0	14733363	male	68	190.0	94.0	29.0	105.0	40.8
1	14861698	female	20	166.0	60.0	14.0	94.0	40.3
2	11179863	male	69	179.0	79.0	5.0	88.0	38.7
3	16180408	female	34	179.0	71.0	13.0	100.0	40.5
4	17771927	female	27	154.0	58.0	10.0	81.0	39.8
5	15130815	female	36	151.0	50.0	23.0	96.0	40.7
6	19602372	female	33	158.0	56.0	22.0	95.0	40.5
7	11117088	male	41	175.0	85.0	25.0	100.0	40.7
8	12132339	male	60	186.0	94.0	21.0	97.0	40.4
9	17964668	female	26	146.0	51.0	16.0	90.0	40.2

Fig2:Calories Data with 7 input attributes

B. *Preprocessing of Dataset:*

Before diving into analysis, it's essential to clean up the data. Outliers, those oddball data points, can mess up our results. Here used a simple method called the Z-score to find and remove them.

1.How Z-score works.

Think of the Z-score as a way to measure how far a data point is from the average. If it's too far away, we consider it an outlier.

Here's the formula: $Z=\sigma x-\mu$.

Here, x is the data point, μ is the average, and σ is how much the data points vary from the average.

2.Cleaning up the data

Once we've calculated the Z-scores for all data points, we look for those with Z-scores that are really high or really low, usually more than 3 or less than -3. These are the outliers we want to get rid of.

3.Visualising The Changes

To show how the data changes after cleaning out the outliers, below mentioned in figure3 and figure4.

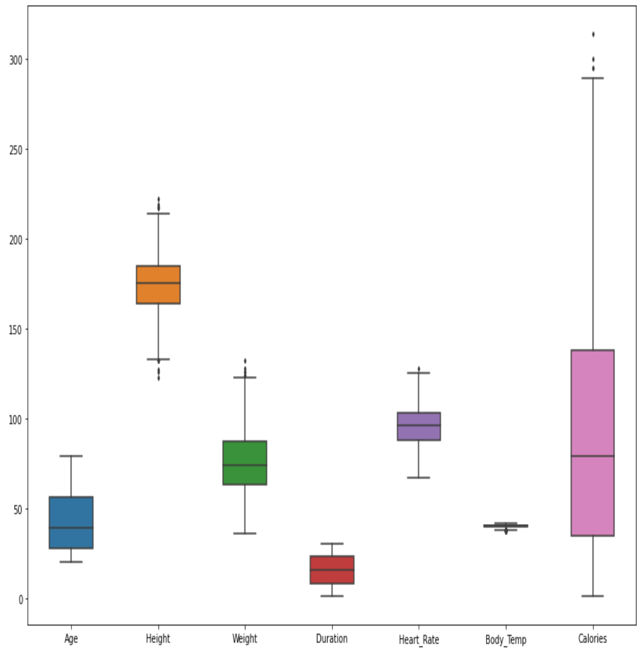


Fig3:Data distribution before outlier removal

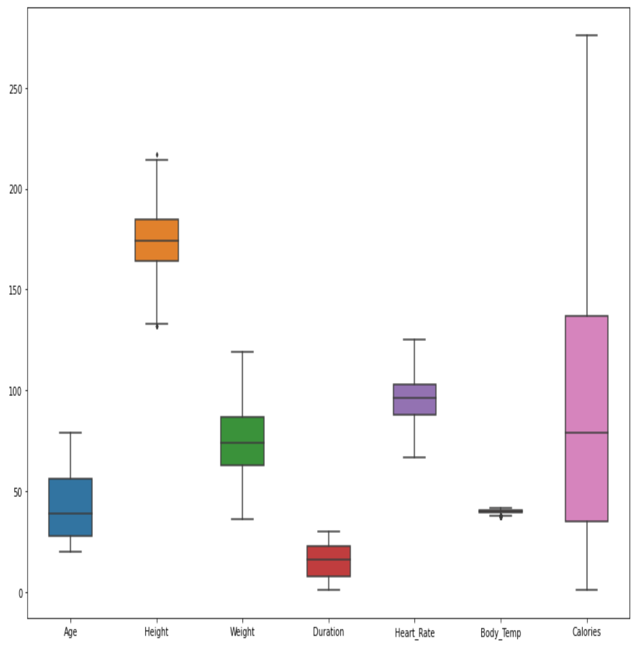


Fig4:Data distribution after outlier removal

C. *Data Analysing and Correlation:*

Prior to predictive modelling, it's essential to analyse the dataset's variables. Key factors like heart rate, exercise duration, and body temperature show strong correlation with calories burned. Additionally, height and weight exhibit notable associations with calorie expenditure. Understanding these relationships lays the foundation for accurate predictive models in estimating calorie burn during physical activities.

According to the data analysis, the feature ID was omitted as it was found to have no influence on predicting an individual's calorie expenditure.

The process of picking out the most important attributes helped us understand which factors are vital for predicting the outcome accurately. By doing this we gained valuable insights and making our predictions more reliable and useful.

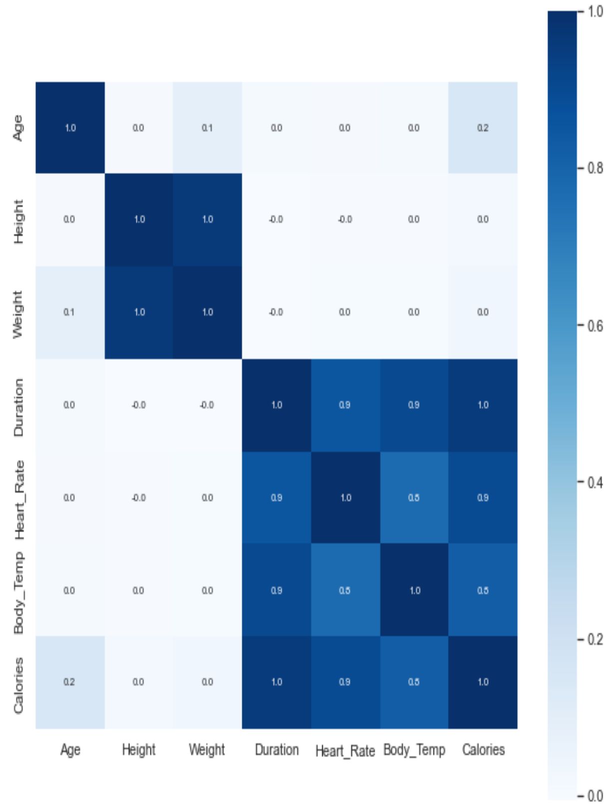


Fig5: Correlation of the Data

D. *Data Visualization:*

We used graphs and charts to look at the data and see if there were any patterns. In below figure we divided the dataset by gender (male and female) on one side and how many records there were on the other.

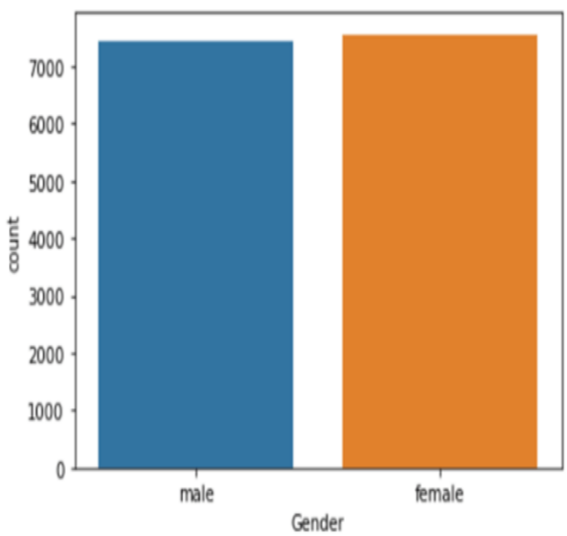


Fig6: Finding the distribution of Gender Column

This visualization allows us to observe how the data is distributed across genders, providing initial insights into potential gender-based differences in the dataset.

To further explore the relationships between key attributes and calorie expenditure, we created plots showcasing the correlations between heart rate, exercise duration, body temperature, and calorie burn.

1. Exercise Duration vs. Calories Burned:

Figure7 illustrates the relationship between exercise duration (in minutes) and calories burned. As duration increases, there is a corresponding rise in calorie expenditure, indicating a positive correlation between the two variables.

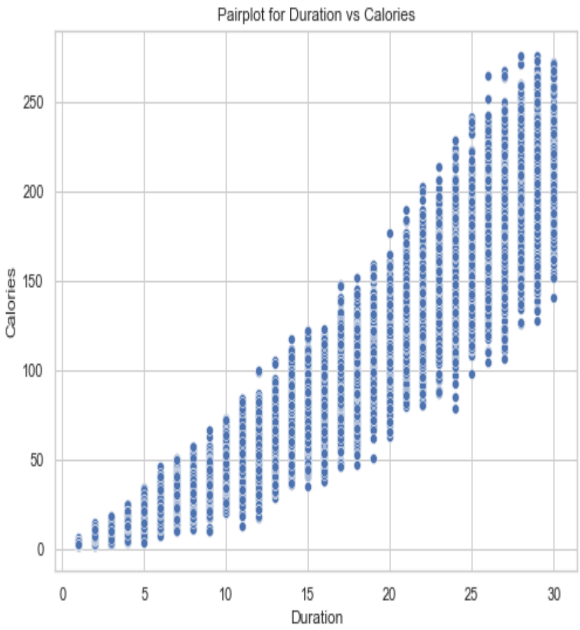


Fig7: Exercise Duration vs Calories Burned

2. Heart Rate vs. Calories Burned:

Figure 8 demonstrates the impact of heart rate (in beats per minute) on calorie burn during physical activities. Higher heart rates are associated with increased calorie burning, indicating a positive relationship between heart rate and calorie expenditure.

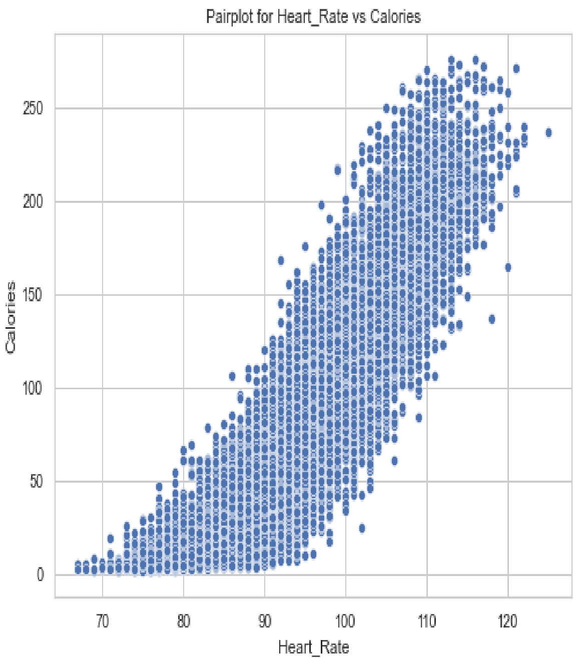


Fig8: Heart Rate vs Calories Burned

3. Body Temperature vs. Calories Burned:

Figure 9 showcases the relationship between body temperature and calories burned. As body temperature increases, there is a corresponding increase in calorie expenditure, suggesting a positive correlation between body temperature and calorie burn.

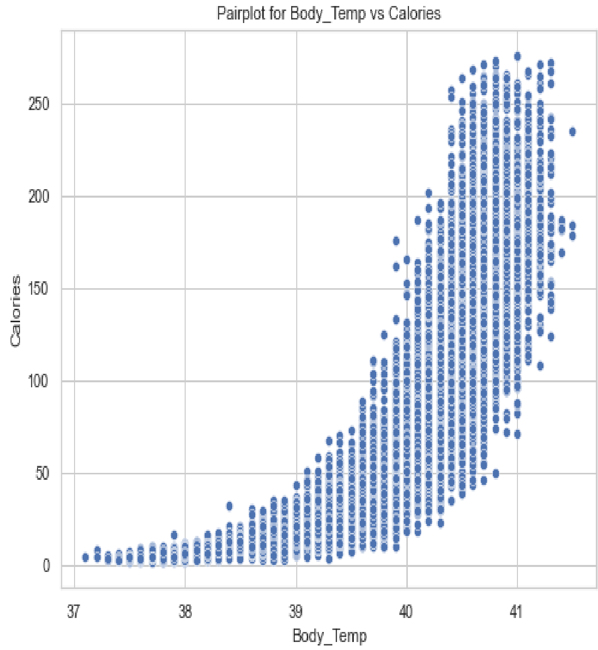


Fig9:Body_temp vs Calories.

E. Data Preparation for Calorie Expenditure Prediction

1. Consideration of All Remaining Attributes

All attributes, except for the feature ID, are retained for subsequent processes. This comprehensive approach ensures that all potential predictors are included, maximizing the information available for developing accurate predictive models of calorie expenditure during physical activities.

2. Label Encoding for Textual Data

In this step, textual data such as gender (male/female) is converted into a numeric format using label encoding. This ensures compatibility with regression algorithms by assigning unique numeric labels to categorical variables, facilitating subsequent analysis.

```
cal_data.replace({'Gender':{'male':0,'female':1}},inplace=True)
```

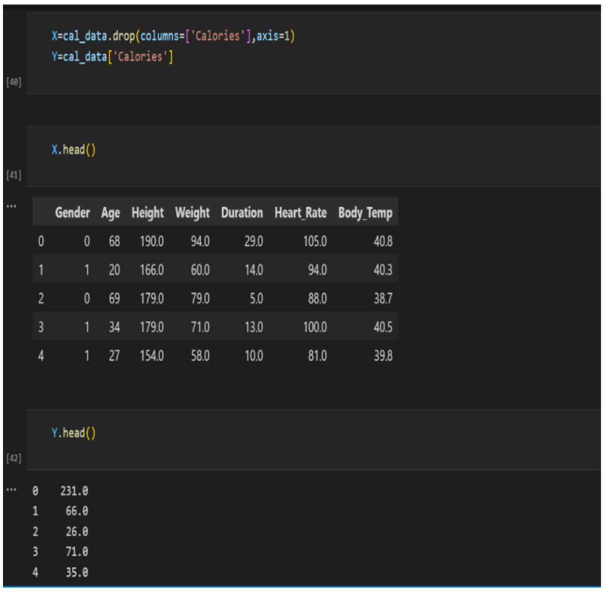
```
cal_data.head()
```

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

Fig10: Text to Numeric Conversion of Gender attribute.

3. Partitioning the Dataset

The dataset is divided into two categories - feature variables and the target variable. Feature variables encompass attributes like gender, age, height, weight, duration, heart rate, and body temperature, while the target variable, "Calories Burned," is isolated for prediction purposes.



```
X=cal_data.drop(columns=['Calories'],axis=1)
Y=cal_data['Calories']
```
[40]

```
X.head()
```
[41]

	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp
0	0	68	190.0	94.0	29.0	105.0	40.8
1	1	20	166.0	60.0	14.0	94.0	40.3
2	0	69	179.0	79.0	5.0	88.0	38.7
3	1	34	179.0	71.0	13.0	100.0	40.5
4	1	27	154.0	58.0	10.0	81.0	39.8

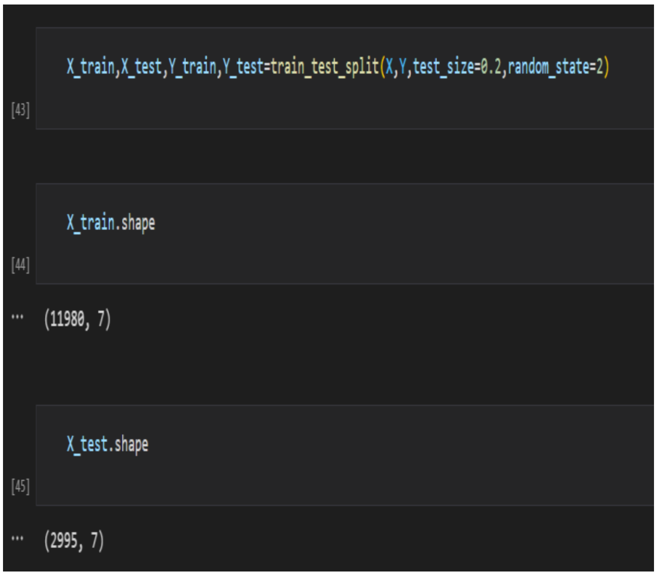
```
Y.head()
```
[42]

```
0    231.0
1     66.0
2     26.0
3     71.0
4     35.0
```

Fig11: Partitioning the dataset into feature and target.

F. Separating and Splitting the Data:

Separating the input features and target data. Then, we split the data into two parts: one part, which was 80% of the data, was used for training our models, and the other part, which was 20% of the data, was used to see how well the models worked.



```
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,random_state=2)
```
[43]

```
X_train.shape
```
[44]

```
(11980, 7)
```

```
X_test.shape
```
[45]

```
(2995, 7)
```

Fig12: Separating the data into train and test

G. Model Assessment and Selection:

By using K-Fold Validation [13] we tested four different machine learning models —Ridge, random forest, linear regression, and XGBoost regression—to see how well they could predict calorie burn.

After conducting testing, the models will undergo evaluation based on their scores obtained through 10 K-fold validation iterations.

The model with the highest score in predicting the target variable will be selected as the best model. Additionally, forecast performance metrics such as mean square error (MSE), root mean square Error (RMSE), and mean absolute error (MAE) will be further considered to evaluate performance standards. The model exhibiting the lowest prediction errors will be deemed the most suitable choice.

H, Building Web App:

This web application gets seven key attributes as input, including heart rate, exercise duration, body temperature, height, weight, and gender, to accurately predict calorie expenditure. Through rigorous testing and evaluation, we determine the most effective predictive model based on metrics such as MSE, RMSE, and MAE, ensuring reliable calorie burn predictions for users.

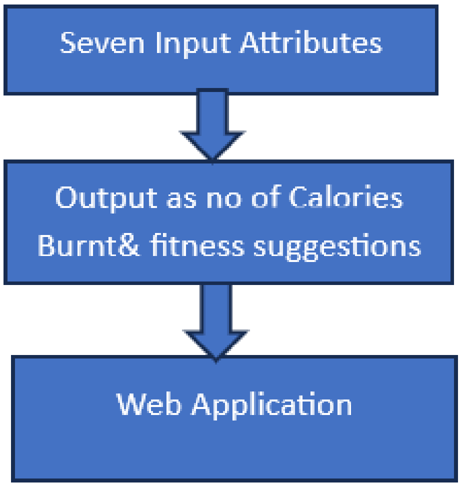


Fig13: Workflow of the Web App

I. Evaluation Metrics of Different Algorithms:

In our examination of various regression methods for predicting calorie expenditure, we rely on a set of essential evaluation measures to assess their effectiveness. These metrics offer numerical benchmarks to gauge how well each model predicts calorie burn based on the given data. Here's an explanation of the metrics we use:

Mean Absolute Error (MAE):

MAE calculates the average difference between predicted and actual values, regardless of direction. It gives us a simple way to understand how far off our predictions are, with lower MAE values suggesting better performance.

Mean Squared Error (MSE):

MSE finds the average of the squared differences between predicted and actual values. By squaring the differences, MSE emphasizes larger errors more than smaller ones. Lower MSE values indicate closer agreement between predicted and actual values.

Root Mean Squared Error (RMSE):

RMSE is the square root of MSE and provides a straightforward measure of prediction accuracy [14].Since it's in the same units as the target variable, it's easy to interpret. Like MAE and MSE, lower RMSE values indicate better model performance.

In evaluating the execution of various regression algorithms, we observe significant variations in their predictive accuracy. Random Forest and XGBoost models outperform Linear and Ridge Regression, demonstrating substantially lower mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) values. Specifically, XGBoost exhibits the lowest errors across all metrics, indicating its superior predictive capability. These findings underscore the importance of selecting the most suitable algorithm for accurate calorie burn prediction. The Bar chart visually compares the metrics above mentioned among different regression models.

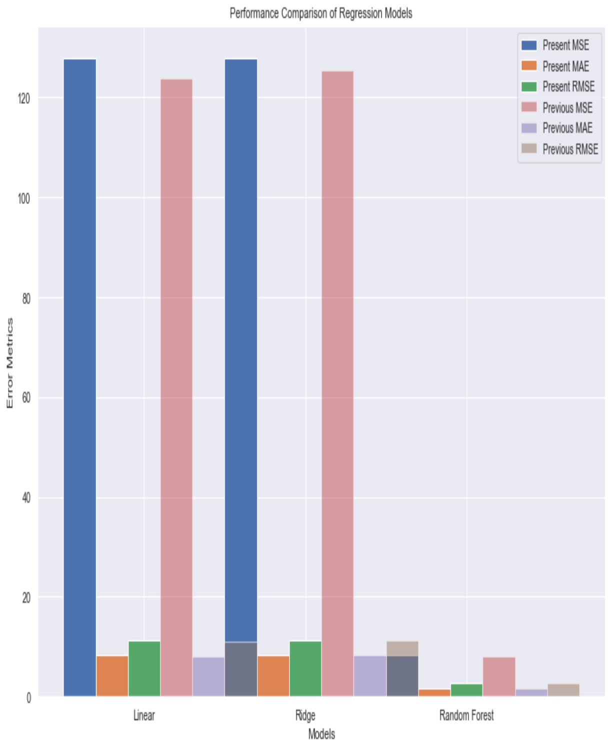


Fig14: BAR chart of evaluation metrics of different algorithms

The table below offers a comprehensive comparison of performance metrics for three distinct regression models. By examining the data provided, researchers can make informed decisions regarding model selection and optimization strategies tailored to the specific needs of their analysis.

Table 1: Model performance of regression model

Algorithm	MAE	MSE	RMSE
Linear	8.36	127.96	11.31
Random Forest	1.78	8. 31	2.88
Ridge	8.36	127.96	11.31
XGBoost	1.43	4.32	2.07

J. Model Accuracies of Different Models:

We further assessed the performance of the model using 10 iterations of K-fold cross-validation, a robust technique commonly employed in machine learning evaluations. These findings are particularly pertinent for integrating calorie prediction into practical applications such as web apps for fitness tracking and health monitoring. The results demonstrate consistently high accuracies across all models. Specifically, Linear Regression [15] achieved an average accuracy of 96.76%, Random Forest performed even better with an average accuracy of 99.81%, while XGBoost exhibited the highest average accuracy of 99.89%. These results underline the reliability and effectiveness of these models in predicting calorie expenditure, making them well-suited for implementation in web applications. The bar graph below illustrates the average accuracies of the regression algorithms. accuracies of the regression algorithms.

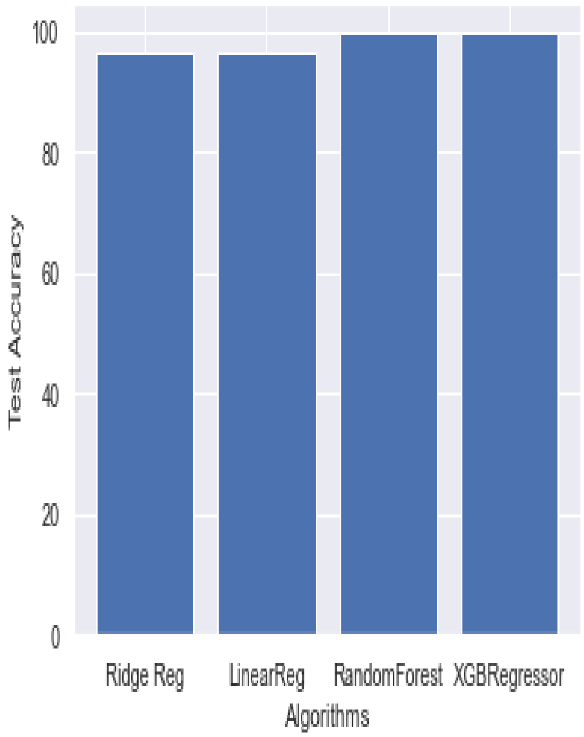


Fig15: Accuracies of Different Regression Models.

IV. Results and Discussion

In the results section, we will analyse the accuracy of models trained on both past and current data, examine various types of errors encountered, present bar charts illustrating the accuracy of different algorithms, and provide visual representations of evaluation metrics for each algorithm.

In below figure the bar chart contrasts the accuracies of four regression models, showcasing their performance in previous instances versus the present. Each bar represents the accuracy of a specific model, allowing for a direct comparison between past and current predictive capabilities.

Additionally, we will showcase the predicted results from our web application and compare our findings to recent studies in the field.

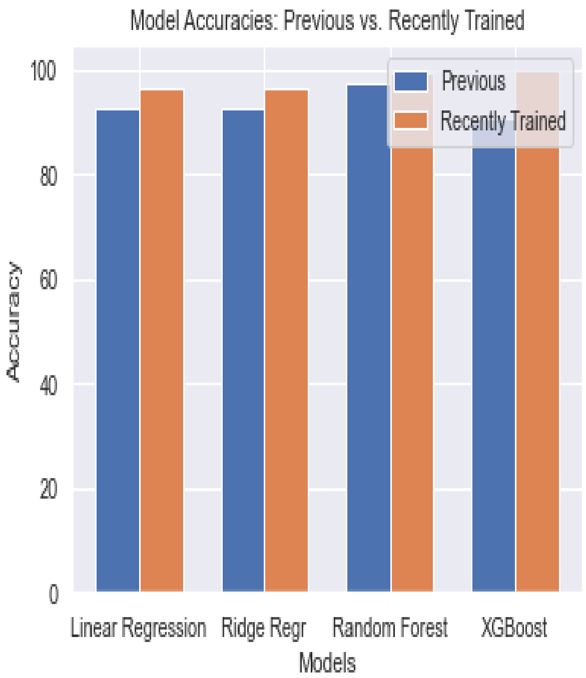


Fig16:Model Accuracies of previous vs present

In figure17, the input attributes form for predicting calories burnt provides users with an interface to input relevant data such as activity duration and intensity level. This allows for accurate estimation of calorie expenditure, enabling informed exercise planning.

In Figure 18, users are presented with the number of calories burnt during the selected activity, accompanied by tailored fitness suggestions. These suggestions serve to optimize the user's workout routine based on their calorie expenditure, aiding in achieving fitness goals effectively.

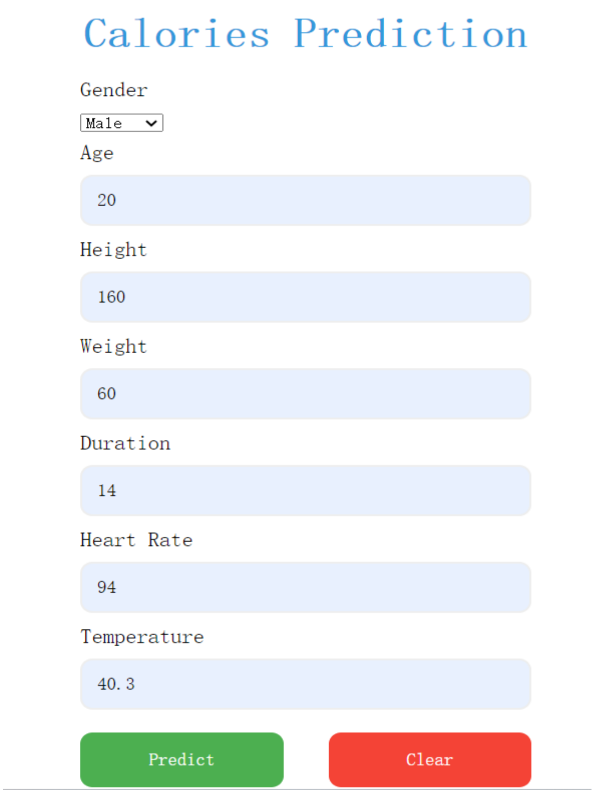


Fig17: Calories Burnt Prediction Form

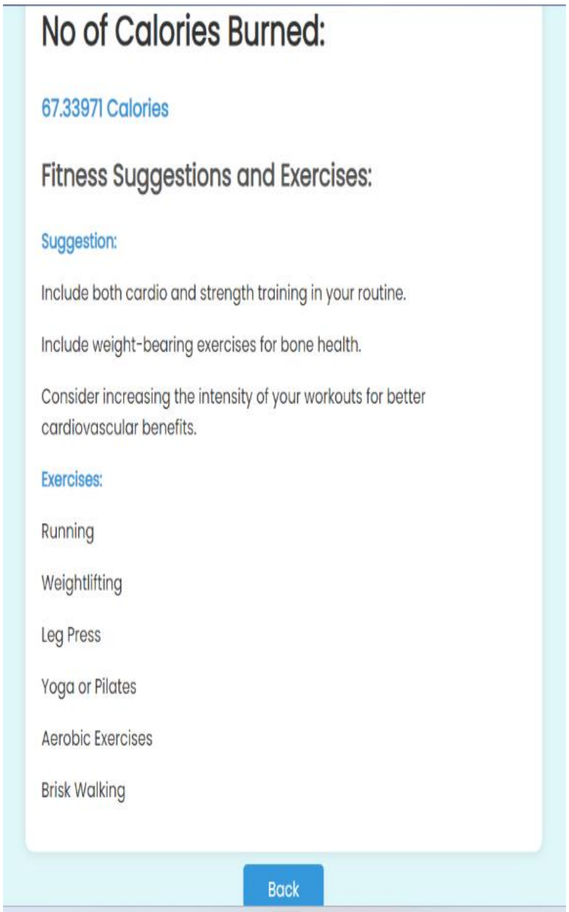


Fig18: Predicted Calories and Suggestions

V. CONCLUSION

After comparing different regression algorithms for predicting calorie expenditure, it's clear that XGBoost predicts more accurately compared to other regression models.

When integrated into web applications for fitness tracking and health monitoring, XGBoost holds promise in providing precise estimations of calorie burn.

Table2: Different Model Accuracies of Existed vs Proposed.

Existed System	Accuracy	Proposed System	Accuracy
Linear	92.88%	Linear	96.75%
Ridge	92.82%	Ridge	96.78%
Random	95.77%	Random	99.72%

XGBoost	-	XGBoost	99.88%

The study also highlights areas for further exploration, such as adjusting parameters and exploring alternative machine learning techniques. Improving the accuracy of the calorie burn measurement model through these means could significantly enhance their usefulness in real-world applications.

In summary, XGBoost is best regression algorithm used for accurately predicting calorie expenditure, with potential implications for advancing health and fitness monitoring through machine learning approaches.

REFERENCES:

[1] 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.

[2] 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.

[3] 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 (WI Mob), 2020, pp. 1-6, doi: 10.1109/WiMob50308.2020.9253424.

[4] 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

[5] 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.

[6] S. T and V. K, "PREDICTION OF USER'S CALORIE ROUTINE USING CONVOLUTIONAL NEURAL NETWORK," International Journal of Engineering Applied Sciences and Technology, vol. 5, no. 3, pp. 189-195, 2020.

[7] S. P. Vinoy and B. Joseph, "Calorie Burn Prediction Analysis Using XGBoost Regressor and Linear Regression Algorithms," in Proceedings of the National Conference on Emerging Computer Applications (NCECA), Kottayam, 2022.

[8] S. S. Ratnakar and V. S, "Calorie Burn Prediction using Machine Learning," International Advanced Research Journal in Science, Engineering and Technology, vol. 9, no. 6, pp. 781-787, June 2022.

[9] R. K. Singh and V. Gupta, "Calories Burnt Prediction Using Machine Learning, "International Journal of Advanced Research in Computer and Communication Engineering, vol. 11, no. 5, May 2022.

[10] M. Nipas, A. G. Acoba, J. N. Mindoro, M. A. F. Malbog, J. A. B. Susa and J. S. Gulmatico, "Burned Calories Prediction using Supervised Machine Learning: Regression Algorithm," in 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, March 2022.

[11] R. S. Biyani and M. S. Nandini, "CALORIES PREDICTION BASED ON FOOD IMAGES," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 8, pp. 2122-2125, August 2020.

[12] K.Westerterp, "Control of energy expenditure in humans," European Journal of Clinical Nutrition, vol. 71, pp. 340-344, 30 November 2016.

[13] Sireesha Moturi, S.N.Tirumala Rao, Srikanth Vemuru,Grey wolf assisted dragonfly-based weighted rule generation for predicting heart disease and breast cancer, Computerized Medical Imaging and Graphics, Volume 91, 2021, 101936, ISSN 0895-6111.

[14] A. Kadam, A. Shrivastava, S. K. Pawar, V. H. Patil, J. Michaelson and A. Singh, "Calories Burned Prediction Using Machine Learning," *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, Gautam Buddha Nagar, India, 2023, pp. 1712-1717,doi:*10.1109/IC3I59117.2023.10397623*.

[15] 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.

[16] 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.