International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

AI-Based Fake Media Detection Using Machine Learning and Natural Language Processing

NAKKA DAVID RAJU

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

A. Durga Devi

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

In recent years, maintaining a healthy lifestyle has become a major concern due to increasing sedentary habits and lifestyle-related diseases. Monitoring calorie expenditure during physical activities plays a vital role in fitness management, weight control, and overall well-being. This project presents an Intelligent Calorie Burn Prediction System that utilizes machine learning techniques to estimate the number of calories burned during various physical activities.The system is developed using Python and integrates a graphical user interface (GUI) built with Tkinter to provide an interactive user experience. It incorporates the XGBoost regression algorithm, a powerful ensemble learning technique known for its high accuracy and efficiency in predictive modeling tasks. The system predicts calorie expenditure based on multiple physiological and activity-related parameters, including age, gender, height, weight, workout type, exercise duration, heart rate, and body temperature.A synthetic dataset of 10,000 records is generated to simulate real-world conditions. The dataset includes diverse combinations of user attributes and activity patterns, ensuring robustness in model training. The categorical variables such as gender and workout type are encoded using Label Encoding, and the dataset is split into training and testing sets for model evaluation. The performance of the model is assessed using the R^2 (coefficient of determination) metric, providing a measure of prediction accuracy.The system also includes a BMI (Body Mass Index) calculator, allowing users to assess their health status. Based on user input, the system predicts calorie burn and visually represents the results using graphical plots. Additionally, dataset visualization helps users understand the relationship between exercise duration and calorie expenditure.This project demonstrates the practical application of machine learning in fitness and health monitoring. It provides a user-friendly platform for individuals to estimate calorie burn and make informed decisions about their physical activities. The system can be further enhanced by integrating real-world datasets, wearable device data, and mobile applications.

Keywords: Calorie Prediction, XGBoost, Machine Learning, Fitness Monitoring, BMI, Regression Model, Health Analytics, Data Visualization

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

I. INTRODUCTION

The growing awareness of health and fitness has led to increased interest in monitoring physical activities and calorie expenditure. Calories burned during exercise is a key indicator of energy expenditure and plays a crucial role in weight management and overall health. Traditional methods of estimating calorie burn often rely on generalized formulas or fitness devices, which may not always provide accurate results due to individual differences.With the advancement of machine learning, it is now possible to develop predictive models that can estimate calorie burn more accurately by considering multiple physiological and environmental factors. Machine learning algorithms can analyze complex relationships between variables such as age, weight, heart rate, and exercise duration, providing personalized predictions.This project focuses on developing an Intelligent Calorie Burn Prediction System using the XGBoost algorithm. XGBoost is a gradient boosting technique that combines multiple weak learners to create a strong predictive model. It is widely used in regression and classification tasks due to its high performance and scalability.The system is designed with a user-friendly graphical interface that allows users to input their personal and activity-related details. Based on this input, the model predicts the number of calories burned. The inclusion of a BMI calculator further enhances the system by providing users with insights into their health status.One of the unique aspects of this project is the use of a synthetic dataset generated programmatically. This approach allows for the creation of a large and diverse dataset, which improves the model's ability to generalize. The dataset includes various combinations of user attributes and activity parameters, making the model robust and reliable.The system also provides graphical visualization of data, helping users understand the relationship between exercise duration and calorie burn. This feature enhances user engagement and provides valuable insights.Overall, this project demonstrates how machine learning can be applied in the domain of fitness and health monitoring. It offers a practical solution for estimating calorie expenditure and encourages users to adopt healthier lifestyles.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

Calorie burn prediction and fitness tracking have been widely studied in recent years, with various approaches ranging from traditional statistical models to advanced machine learning techniques. Early methods relied on simple mathematical formulas based on factors such as weight, age, and activity type. While these methods provided basic

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

estimates, they lacked accuracy due to their inability to capture complex relationships among variables.With the emergence of machine learning, researchers have explored various algorithms for predicting calorie expenditure. Linear regression models were initially used due to their simplicity and interpretability. However, these models often failed to capture nonlinear relationships present in physiological data.Support Vector Machines (SVM) and Decision Trees have been applied to improve prediction accuracy. Decision Trees provide better interpretability, while SVMs handle high-dimensional data effectively. However, these methods still have limitations in handling large datasets and complex interactions.

Ensemble learning techniques such as Random Forest and Gradient Boosting have shown significant improvements in prediction performance. These methods combine multiple models to reduce variance and bias, resulting in more accurate predictions. XGBoost, an optimized implementation of gradient boosting, has gained popularity due to its efficiency and scalability.Recent studies have also explored deep learning approaches, including Artificial Neural Networks (ANNs), for calorie prediction. These models can capture highly complex patterns but require large datasets and computational resources.Wearable devices and IoT-based fitness trackers have also contributed to this field by providing real-time data such as heart rate and activity levels. Integrating such data with machine learning models has improved prediction accuracy. However, these systems often require expensive hardware and infrastructure.The proposed system builds upon these advancements by using XGBoost for accurate calorie prediction and integrating it into a user-friendly GUI application. Unlike many existing systems, it does not rely on external devices and provides an accessible solution for users.

III. EXISTING SYSTEM

Existing calorie prediction systems primarily rely on fitness trackers, mobile applications, and basic estimation formulas. These systems typically use predefined equations that consider limited parameters such as weight, duration, and activity type. While they provide quick estimates, their accuracy is often compromised due to the lack of personalization.Fitness tracking devices such as smartwatches and wearable sensors have improved calorie estimation by incorporating real-time data such as heart rate and movement. However, these devices are often expensive and may not be accessible to all users. Additionally, their accuracy depends on sensor quality and calibration.Many mobile applications also provide calorie tracking features. These apps use built-in algorithms and user inputs to estimate calorie burn. However, they often rely on generic models that do not adapt to individual variations, leading to less accurate results.Traditional statistical models used in these systems are limited in their ability to capture complex relationships between multiple variables. They often fail to consider factors such as body temperature, workout intensity, and individual physiological

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

differences.Another limitation of existing systems is the lack of visualization and user interaction. Most applications provide only numerical outputs without offering insights into the underlying data or relationships.The proposed system addresses these limitations by using a machine learning approach, specifically XGBoost, to provide more accurate and personalized predictions. It also includes data visualization and an interactive GUI, making it more user-friendly and informative.

IV. PROPOSED METHOD

The proposed system is an **Intelligent Calorie Burn Prediction System** that leverages machine learning, specifically the XGBoost regression algorithm, to accurately estimate calories burned during physical activities. Unlike traditional systems that rely on static formulas, this system dynamically analyzes multiple physiological and activity-based parameters to deliver personalized predictions.The system is composed of three main modules: **dataset generation module, machine learning prediction module, and user interaction module (GUI)**. A synthetic dataset is generated with 10,000 records containing attributes such as gender, age, height, weight, workout type, duration, heart rate, and body temperature. These features are crucial in determining energy expenditure and are widely used in recent research for calorie prediction .The prediction module uses the XGBoost regressor, which is known for its high accuracy and ability to model nonlinear relationships. The model is trained using encoded and preprocessed data, and its performance is evaluated using the R^2 score. Studies show that XGBoost outperforms traditional regression models due to its gradient boosting framework and regularization techniques .The system also includes a **BMI calculator** to provide additional health insights. A graphical interface allows users to input their data, view predictions, and visualize results through plots. This enhances usability and user engagement.The proposed system addresses key limitations of existing approaches by providing **personalized, accurate, and interactive calorie predictions**. It can be extended with real-time wearable device data, as modern research highlights the importance of integrating real-time activity metrics for improved prediction accuracy .

V. IMPLEMENTATION

The system is implemented using Python, combining machine learning libraries with a graphical user interface to create a complete application.

1. Dataset Generation

A synthetic dataset is generated using Python libraries such as NumPy and Pandas. The dataset consists of 10,000 records with features including:

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

- Gender
- Age
- Height and Weight
- Workout type
- Duration
- Heart rate
- Body temperature

The calorie value is calculated using a predefined formula that simulates real-world energy expenditure patterns. This approach ensures sufficient data diversity for model training.

2. Data Preprocessing

Categorical variables such as gender and workout type are converted into numerical form using Label Encoding. The dataset is then split into training and testing sets using an 80:20 ratio.

3. Model Training

The system uses the **XGBoost Regressor**, configured with parameters such as:

- Number of estimators
- Learning rate
- Maximum tree depth

The model is trained on the training dataset and evaluated using the R^2 score. Research indicates that XGBoost achieves high prediction accuracy due to its ability to handle complex and nonlinear relationships .

4. Prediction Module

User inputs are collected through the GUI. The system:

1. Encodes categorical inputs
2. Forms a feature array
3. Passes it to the trained model
4. Displays predicted calorie output

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

5. GUI Implementation

The graphical user interface is built using Tkinter and includes:

- Dataset generation tab
- Model training tab
- BMI calculator tab
- Prediction tab

The interface ensures ease of use and accessibility for non-technical users.

6. Visualization

Matplotlib is used to generate:

- Scatter plots for dataset analysis
- Individual prediction graphs

7. Error Handling

The system includes exception handling for:

- Missing dataset
- Invalid inputs ,Untrained model

VI. **ALGORITHMS:**The primary algorithm used in this system is **XGBoost (Extreme Gradient Boosting)**, a powerful ensemble learning method.

1. XGBoost Algorithm

XGBoost builds multiple decision trees sequentially, where each new tree corrects the errors of previous ones. It minimizes a loss function using gradient descent and includes regularization to prevent overfitting.

Key advantages:

- Handles nonlinear relationships
- High accuracy
- Efficient computation

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

XGBoost is widely recognized for its scalability and performance in regression tasks .

2. Label Encoding

Categorical features such as gender and workout type are converted into numeric values to make them compatible with the model.

3. Train-Test Split

The dataset is divided into training and testing sets to evaluate model performance and avoid overfitting.

4. Evaluation Metric (R^2 Score)

The R^2 score measures how well the predicted values match actual values. A higher R^2 indicates better model performance.

5. Working Process

1. Input data collection
2. Data preprocessing
3. Model training
4. Prediction
5. Result visualization

Machine learning models like XGBoost and Random Forest have been proven effective in calorie prediction tasks due to their ability to handle complex datasets .

VII. SYSTEM DESIGN

The system follows a **modular and layered architecture**, ensuring scalability and maintainability.

1. Architecture Overview

The system consists of three layers:

a) Presentation Layer

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

- Built using Tkinter GUI
- Handles user interaction
- Displays predictions and graphs

b) Application Layer

- Implements business logic
- Handles data processing
- Executes machine learning models

c) Data Layer

- Stores dataset (CSV file)
- Maintains trained model

2. System Workflow

1. User generates dataset
2. Model is trained
3. User inputs data
4. System predicts calories
5. Results are displayed and visualized

3. Data Flow Diagram

Input → Preprocessing → Model → Prediction → Output

4. Design Features

- Modular design
- Easy integration of new models
- Interactive interface ,Real-time prediction

5. Advantages

- Accurate predictions
- User-friendly interface
- Cost-effective solution
- No external hardware required

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

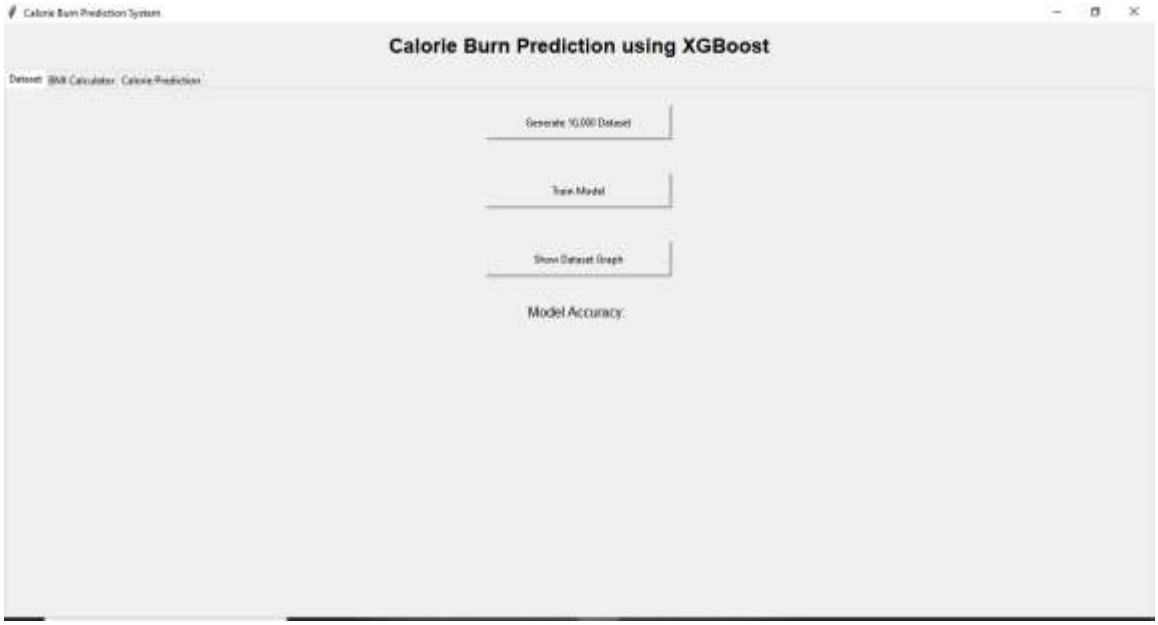
6. Scalability

The system can be extended to:

- Web applications
- Mobile apps
- Integration with wearable devices

Recent studies emphasize the importance of integrating machine learning models into real-time fitness systems for better health monitoring .

SYSTEM DESIGN IMAGES

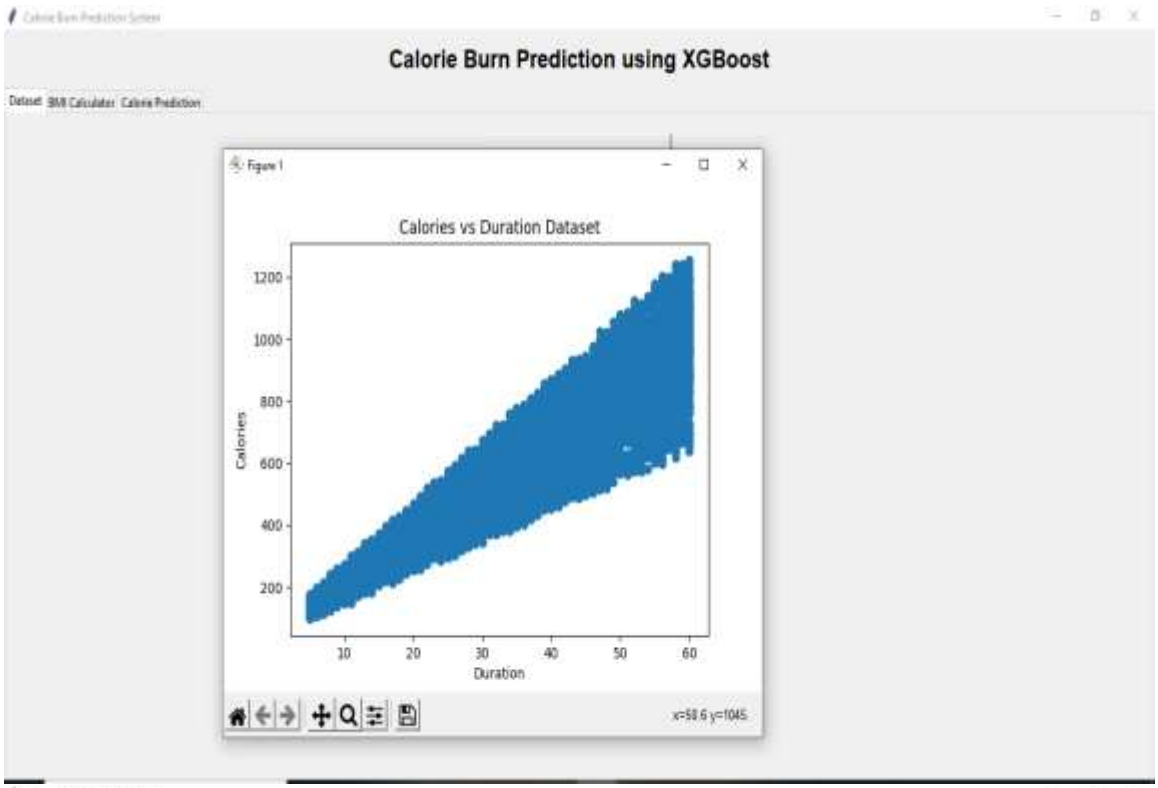


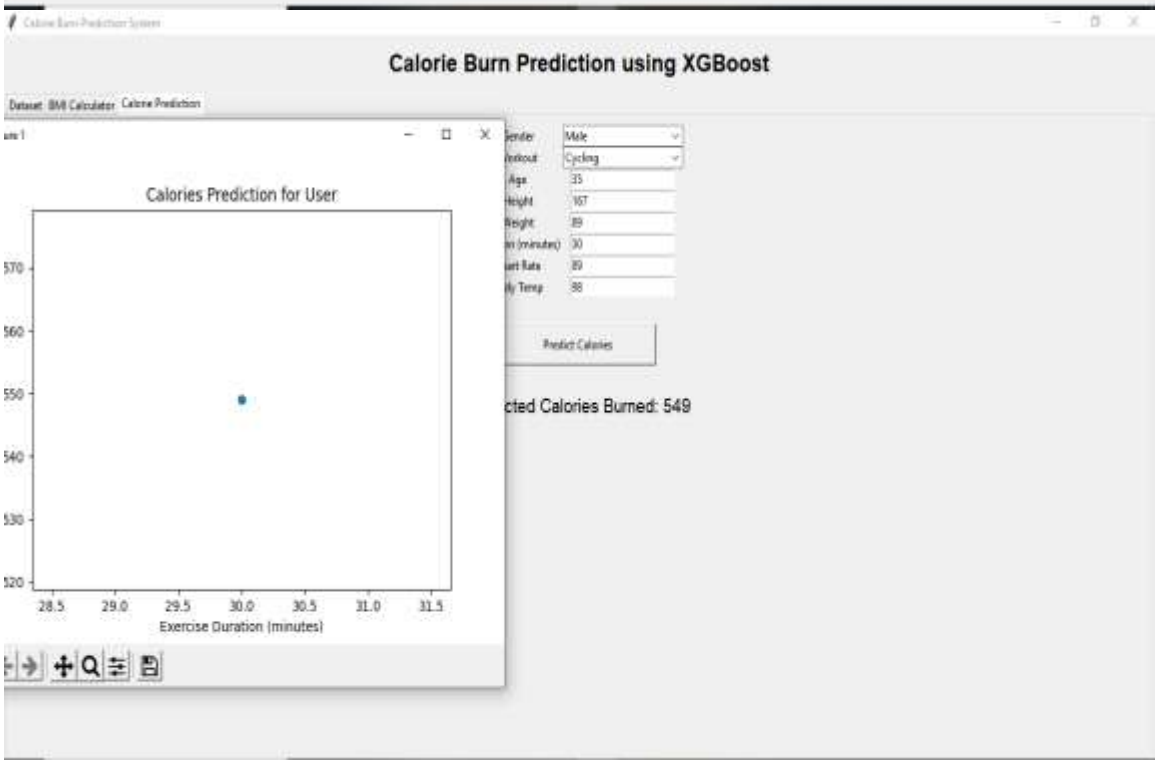
International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper





International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

VIII. CONCLUSION

The Intelligent Calorie Burn Prediction System demonstrates the effective use of machine learning in the domain of fitness and health monitoring. By utilizing the XGBoost algorithm, the system provides accurate and personalized predictions of calorie expenditure based on multiple physiological parameters.The integration of a graphical user interface makes the system accessible to users without technical expertise, enhancing usability and adoption. The inclusion of additional features such as BMI calculation and data visualization further enriches the user experience.Compared to traditional methods, the proposed system offers improved accuracy and adaptability by analyzing complex relationships among variables. Research confirms that machine learning models significantly outperform conventional estimation techniques in predicting calorie burn .The system also provides a scalable framework that can be extended with real-time data from wearable devices, enabling continuous monitoring and more precise predictions. Future improvements may include deep learning models, cloud deployment, and mobile integration.Overall, this project highlights the potential of machine learning in transforming fitness tracking systems and promoting healthier lifestyles.

REFERENCES

1. Sumithra Devi et al. (2025). *Hybrid Model Analysis for Calorie Prediction Using XGBoost and RF*
2. Lin & Zhang (2026). *ML Approach for Calorie Consumption with SHAP Analysis*
3. Aziz et al. (2023). *Calories Burnt Prediction Using Machine Learning*
4. Ratnam & Rao (2025). *Calorie Predictive Activity Analyzer using ML*
5. Bhuiyan et al. (2024). *Workout Analytics Using ML Models*
6. Rayyan et al. (2025). *Calories Burnt Prediction Using ML*
7. Tan et al. (2026). *Comparison of ML Methods for Calories Prediction*
8. Budiman et al. (2024). *XGBoost Optimization for Calorie Prediction*
9. Tejaswini & Rabindranath (2025). *AI Fitness Chatbot with Calorie Prediction*
10. Valmiki & Ambili (2023). *ML Approaches for Calorie Analysis*
11. Ali et al. (2025). *Deep Learning for Calorie Prediction*
12. Chen & Guestrin (2016). *XGBoost: A Scalable Tree Boosting System*
13. Narang (2025). *Multimodal ML for Calorie Estimation*

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

14. Ruede et al. (2020). *Multi-task Learning for Calorie Prediction*
15. Miah et al. (2023). *Mobile Health ML Prediction Systems*