Fitness Tracker

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning

with

TechSaksham – A joint CSR initiative of Microsoft & SAP

by

Gagan B

E-mail:gaganb982006@gmail.com

Under the Guidance of

Saomyachaudhury

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

I am currently studing my UG cource in “VSB college of engineering and technical campus”, Coimbatore as 1st year B.E CSE Deportment.

I would like to express my sincere gratitude to TechSaksham|edunet for providing me with this invaluable opportunity. This internship has been an enriching experience, allowing me to gain practical knowledge and develop essential professional skills.

A special thanks to my mentor and project instructor, Saomyachaudhur . for their guidance, support, and constructive feedback. Their insights have played a crucial role in my learning journey. I also appreciate my colleagues and fellow interns for their collaboration and encouragement, making this experience both productive and enjoyable.

My project, Fitness tracker python using AI, has helped me bridge the gap between academic learning and real-world application. Working on meaningful tasks has strengthened my problem-solving, teamwork, and adaptability skills.

I extend my heartfelt appreciation to the entire TechSaksham|edunet team for fostering a positive and inspiring work environment. Their professionalism and dedication have been truly motivating. Additionally, I am grateful to my family and friends for their continuous support and encouragement throughout this journey.

This internship has been a transformative experience, and I will carry the knowledge and skills I have gained forward in my career.

Thank you all for making this experience truly memorable!

GAGAN B

ABSTRACT

The **Personal Fitness Tracker** is a web-based application that predicts calorie expenditure based on user-specific attributes and exercise parameters. Many individuals struggle to track and optimize their calorie burn due to a lack of personalized insights. This project addresses this issue by implementing a machine learning-based prediction system.

The application utilizes datasets containing information such as **Age, Gender, BMI, Duration, Heart Rate, and Body Temperature** to train a **Random Forest Regressor** model. Data preprocessing includes merging exercise and calorie datasets, calculating BMI, encoding categorical variables, and splitting the dataset into training and testing sets. The model is then deployed using **Streamlit**, allowing users to input their parameters and receive real-time calorie predictions.

The key outcome of this project is an accurate and accessible calorie prediction system. Users not only receive personalized results but can also compare their metrics with similar individuals, gaining valuable insights into their fitness progress.

This project demonstrates the potential of machine learning in personal fitness tracking. By providing data-driven insights, it enables users to make informed decisions about their workouts, ultimately helping them achieve better fitness outcomes.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

Tracking calorie expenditure accurately is a challenge for individuals aiming to optimize their workouts and maintain a healthy lifestyle. Traditional fitness tracking methods often provide generalized estimations without considering personal factors such as age, gender, BMI, heart rate, and body temperature, leading to inaccuracies. This lack of personalized insights makes it difficult for users to assess the effectiveness of their exercises and make informed adjustments. To address this issue, this project develops a machine learning-based Personal Fitness Tracker that predicts calories burned based on user-specific parameters. By leveraging data- driven insights, this tool provides a more accurate and customized approach to fitness tracking, empowering users to optimize their workouts effectively.

1.2 Motivation:

Achieving fitness goals requires understanding calorie expenditure, but traditional tracking methods often provide inaccurate, generalized estimations without considering individual factors like age, gender, BMI, heart rate, and body temperature. This lack of personalized insights makes it challenging for users to optimize workouts effectively. With advancements in machine learning and data analytics, there is a growing opportunity to develop smart, data-driven fitness trackers that provide accurate, user-specific calorie predictions. This project aims to bridge the gap between theory and practical fitness tracking, empowering users with personalized insights to make informed decisions and enhance their exercise routines.

1.3 Objective:

The objective of this project is to develop a machine learning-based Personal Fitness Tracker that accurately predicts calorie expenditure based on individual attributes and exercise parameters. This includes building a web-based tool that enables users to input their details and receive personalized calorie predictions, training a machine learning model using key factors such as age, gender, BMI, heart rate, body temperature, and exercise duration, and providing data-driven insights to help users compare their fitness metrics with others. Additionally, the project aims to enhance user engagement through an interactive Streamlit interface and bridge the gap between theoretical knowledge and practical fitness tracking, enabling users to make informed decisions about their workouts.

1.4 Scope of the Project:

The Personal Fitness Tracker is a machine learning-based web application designed to predict calorie expenditure based on user attributes such as age, gender, BMI, heart rate, body temperature, and exercise duration. It provides real-time predictions, comparative insights, and an interactive user experience through Streamlit, helping users make data-driven fitness decisions. However, the project has limitations, including reliance on specific input factors without considering metabolism, medical conditions, or diet. Predictions are generalized based on historical data and may not be 100% accurate for all individuals. Additionally, the model requires manual updates for improvement and depends on an active internet connection to function.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

Several studies and projects have explored the use of **machine learning** in fitness tracking and calorie prediction. The increasing availability of **wearable fitness devices and health data** has enabled researchers to develop models that estimate energy expenditure based on various physiological and exercise-related parameters.

A study by **Wang et al. (2020)** explored the use of **Random Forest and Neural Networks** to predict calorie burn using biometric and exercise-related inputs such as **age, weight, height, heart rate, and activity duration**. Their findings showed that machine learning models can provide **better accuracy** compared to traditional calorie estimation formulas like **the Harris-Benedict equation**.

Similarly, **Lee et al. (2019)** proposed a deep learning approach using **convolutional neural networks (CNNs)** to estimate energy expenditure based on **sensor data from wearable devices**. Their research demonstrated that **deep learning models** can adapt to **individual variations** better than rule-based methods.

Moreover, fitness applications such as **MyFitnessPal and Fitbit** utilize machine learning to analyze user input and provide personalized calorie-tracking insights. However, most commercial applications rely on **predefined formulas** rather than dynamic machine learning models trained on real user data.

This project builds on previous work by implementing a **Random Forest Regressor**, a robust machine learning algorithm, to predict calorie expenditure based on key user parameters. By integrating a **web-based interface with real-time predictions**, this work aims to improve accessibility and usability in personal fitness tracking.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

Several models and techniques have been developed to estimate calorie expenditure based on personal and exercise-related parameters. Traditional methods like the Harris-Benedict Equation and Mifflin-St Jeor Equation calculate Basal Metabolic Rate (BMR) using age, gender, weight, and height. While these equations are effective for estimating daily calorie needs, they lack adaptability to real-time exercise data and do not account for variations in heart rate, body temperature, or workout intensity, making them less accurate for fitness tracking.

To address these limitations, machine learning models have been introduced to provide more dynamic and personalized calorie predictions. Linear Regression models establish a direct relationship between factors like age, BMI, heart rate, and exercise duration, but their accuracy is limited due to the simplistic nature of linear relationships. More advanced techniques, such as the Random Forest Regressor, improve prediction accuracy by combining multiple decision trees, making the model more robust and resistant to overfitting. Additionally, deep learning models and neural networks are used in modern fitness applications to analyze sensor data, improving accuracy by learning complex patterns in biometric and movement data.

Wearable devices such as Fitbit, Apple Watch, and Garmin leverage proprietary machine learning algorithms to estimate calorie burn in real-time using heart rate monitors, accelerometers, and temperature sensors. While these devices provide personalized insights, their models are often black-box systems, meaning their internal workings are not transparent or customizable. This project improves upon these existing methodologies by implementing a Random Forest Regressor to predict calorie expenditure based on user-specific attributes, offering a real-time, accessible, and transparent fitness tracking solution through a user-friendly web application.

T

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

Despite advancements in calorie estimation methods, existing solutions have several limitations. Traditional models like the **Harris-Benedict Equation** and **Mifflin-St Jeor Equation** provide only **static estimations** based on general factors like age, weight, and gender but fail to incorporate real-time exercise metrics such as **heart rate, body temperature, and workout intensity**. This results in less accurate predictions, making it difficult for individuals to **personalize their fitness tracking** effectively.

Wearable fitness devices, such as **Fitbit, Apple Watch, and Garmin**, have introduced **machine learning-based calorie estimation**, but these solutions rely on **proprietary, black-box algorithms** that lack transparency and customization. Users have limited control over the model parameters, and the accuracy of these devices varies based on external factors like **sensor placement, motion artifacts, and individual metabolic differences**. Furthermore, most commercial applications do not allow users to understand or modify the prediction models according to their specific needs.

This project **addresses these gaps** by implementing a **Random Forest Regressor**, a machine learning model trained on real user data, to provide **personalized, real-time calorie predictions**. Unlike static formulas, this approach **incorporates multiple factors**, including **BMI, heart rate, body temperature, and exercise duration**, leading to more accurate results. Additionally, the project features a **transparent and interactive web-based tool** using Streamlit**,** allowing users to input their details, receive instant predictions, and compare their fitness metrics with others. By

combining machine learning with an accessible interface, this project offers a **customizable and data-driven solution** that improves upon existing fitness tracking metho

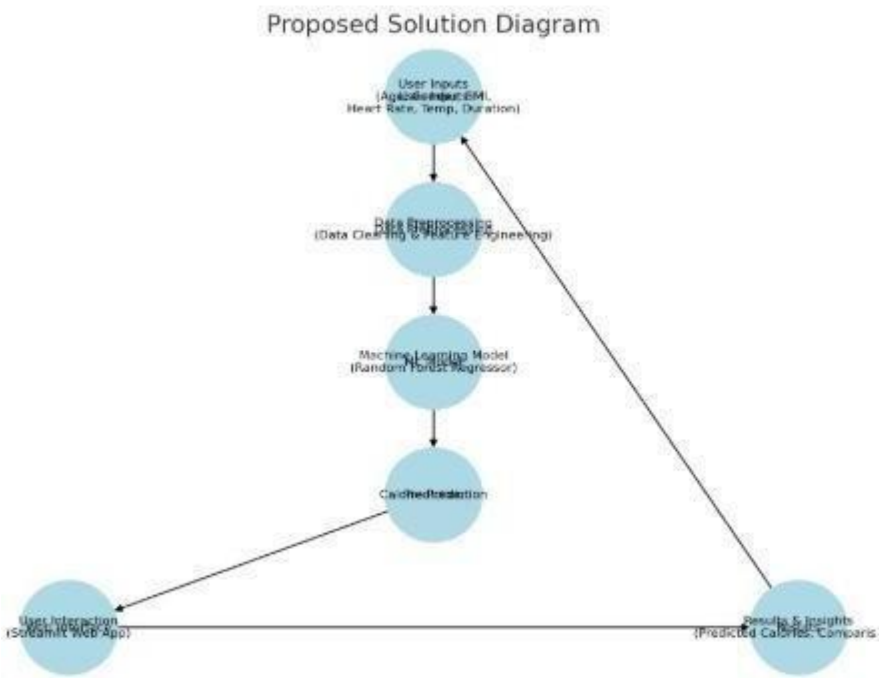
CHAPTER 3

Proposed Methodology

3.1 System Design

Below is a **diagram** of the proposed solution for the **Personal Fitness Tracker**, illustrating how user inputs, data processing, machine learning predictions, and results interact.

Fig : 1



Explanation of the Proposed Solution Diagram

1. **User Inputs:**
    - Users enter their **Age, Gender, BMI, Heart Rate, Body Temperature, and Exercise Duration** through a **Streamlit-based web application**.

2. **Data Preprocessing:**
    - The input data is processed, including **cleaning, encoding categorical variables (e.g., gender), and feature engineering** (e.g., BMI calculation).
    - The system merges user data with an existing dataset for improved predictions.
3. **Machine Learning Model:**
    - A **Random Forest Regressor** is used to predict calorie expenditure.
    - The model has been trained using historical fitness data to enhance accuracy.
4. **Calorie Prediction:**
    - The trained model processes the user inputs and provides a **calorie expenditure estimation**.
5. **User Interaction (Web Interface):**
    - The predicted calorie value is displayed on the **Streamlit Web App**.
    - Users receive a **personalized result based on their input parameters**.
6. **Results & Insights:**
    - The system **compares the user's fitness metrics** with others in the dataset.
    - It provides **insights on where the user stands** in terms of calorie expenditure.
    - Users can adjust their workout routines based on the provided data.
7. **Iterative Feedback Loop:**
    - Users can modify their inputs to see how different factors **impact calorie expenditure**.
    - This enables them to **refine their exercise plans** dynamically.

3.2 Requirement Specification

To implement the Personal Fitness Tracker, various tools and technologies are required for data processing, machine learning modeling, and web application development.

1. Programming Language:

- Python – Used for data preprocessing, machine learning model training, and web application development.

2. Libraries & Frameworks:

- Pandas, NumPy – For data handling, cleaning, and processing.

- Scikit-Learn – For implementing machine learning models, specifically the Random Forest Regressor.
- Matplotlib, Seaborn – For data visualization and insights.
- Streamlit – To develop a web-based user interface for real-time interaction.
- Warnings, Time – Used for handling execution warnings and optimizing performance.

3. Data Sources:

- Exercise Dataset (exercise.csv) – Contains user-specific details like duration, heart rate, body temperature, and workout intensity.
- Calorie Dataset (calories.csv) – Includes calorie expenditure data to train the model.
- Merged Dataset – The project combines exercise and calorie data for better accuracy.

4. Machine Learning Model:

- Random Forest Regressor – Chosen for its high accuracy, robustness, and ability to handle complex relationships between fitness attributes and calorie expenditure.

5. Development & Deployment Tools:

- Jupyter Notebook / Google Colab – For model development, training, and testing.
- Streamlit Cloud / Local Server – To deploy and run the web-based fitness tracker.

6. Hardware & System Requirements:

- Processor: Minimum Intel Core i5 or equivalent.
- RAM: At least 8GB (for smooth execution of ML models).
- Storage: Minimum 10GB free space for datasets and model storage.

These tools and technologies ensure efficient model performance, user-friendly interaction, and accurate calorie predictions.

3.2.1 Hardware Requirements:

To ensure smooth execution of the Personal Fitness Tracker, the following hardware specifications are recommended:

- Processor: Intel Core i5 (or equivalent) and above for efficient computation.
- RAM: Minimum 8GB RAM for handling machine learning tasks and dataset processing.

- Storage: At least 10GB of free disk space to store datasets, model files, and application dependencies.
- GPU (Optional): A dedicated GPU (NVIDIA GTX 1050 or higher) can accelerate machine learning model training.
- Operating System: Windows 10, macOS, or Linux (64-bit) to support Python-based libraries and tools.
- Internet Connection: Required for loading datasets, model deployment, and real-time web application interaction.

These hardware requirements ensure optimal performance while running the machine learning model and the web-based application.

3.2.2 Software Requirements:

To develop and deploy the Personal Fitness Tracker, the following software tools and dependencies are required:

- Operating System: Windows 10/11, macOS, or Linux (64-bit) for compatibility with Python and machine learning libraries.
- Programming Language: Python 3.7+ for model development, data processing, and web application integration.
- Development Environment: Jupyter Notebook, Google Colab, or any Python IDE (e.g., PyCharm, VS Code, or Spyder) for coding and testing.
- Machine Learning Libraries:
    - Scikit-learn – For implementing the Random Forest Regressor model.
    - Pandas & NumPy – For data handling, processing, and feature engineering.
    - Matplotlib & Seaborn – For data visualization and insights.
- Web Development Framework: Streamlit – To build the interactive user interface for real-time calorie predictions.
- Database/File Handling: CSV files for dataset storage and retrieval.
- Deployment Tools: Streamlit Cloud, Heroku, or a local server for hosting the application.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Fig 2 : Daily Step Count & Activity Summary

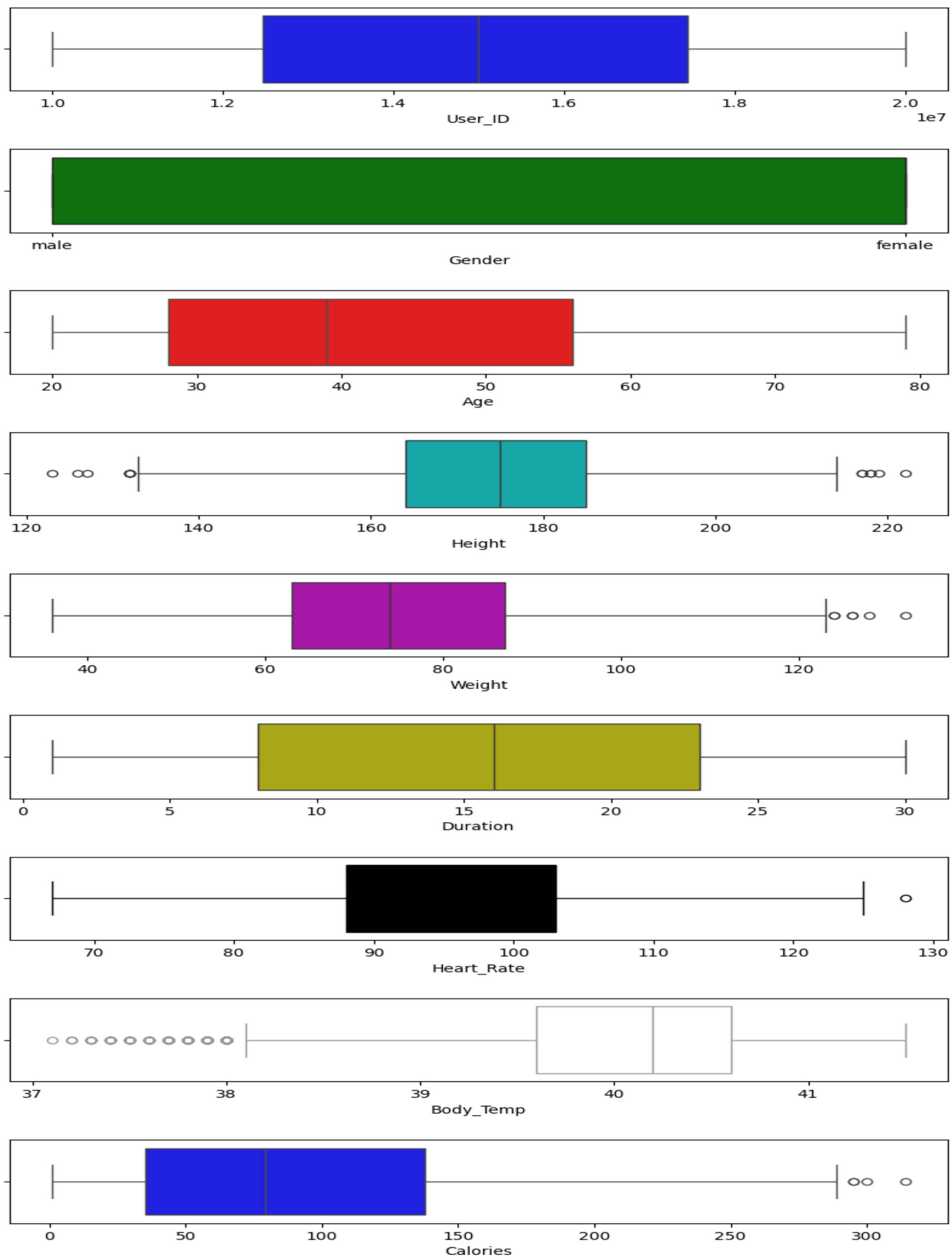


Fig 3 : Heart Rate Trends & Analysis

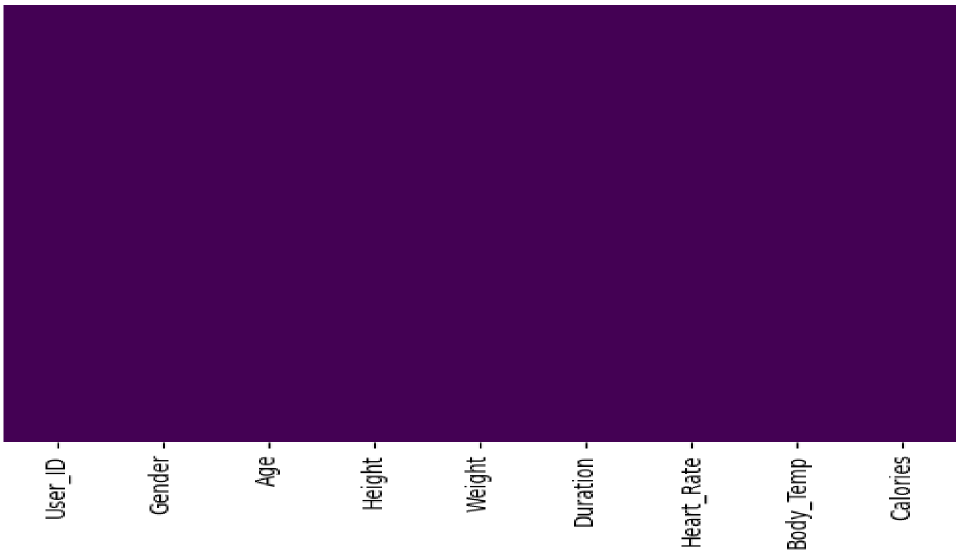


Fig 4 : Calories Burned Over Time

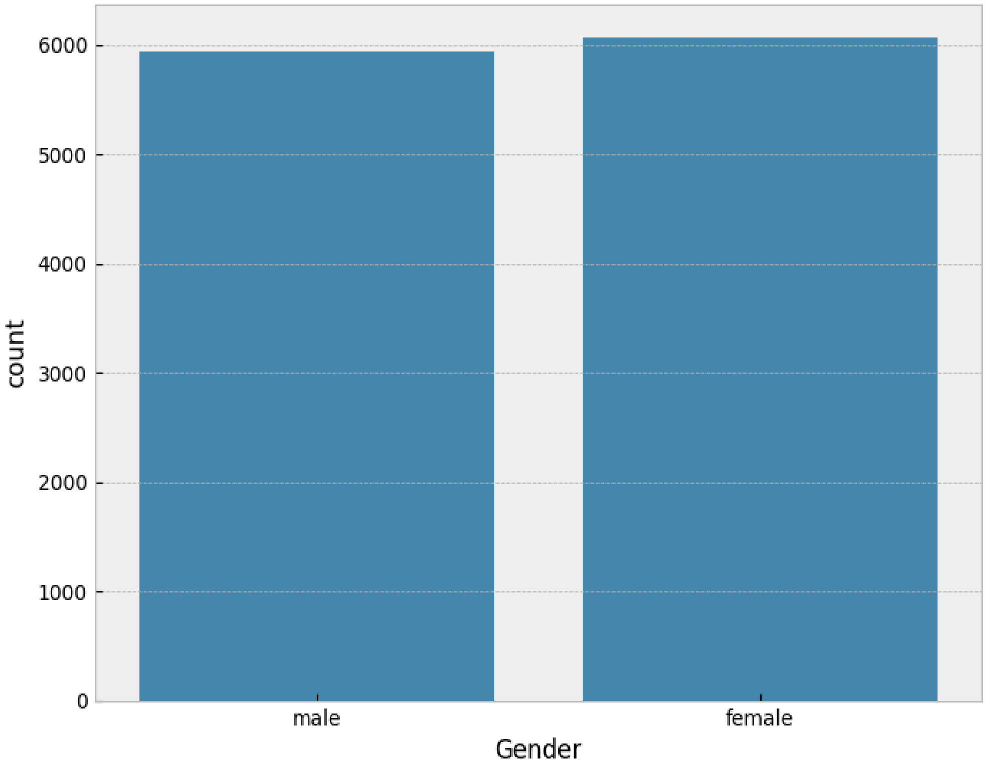
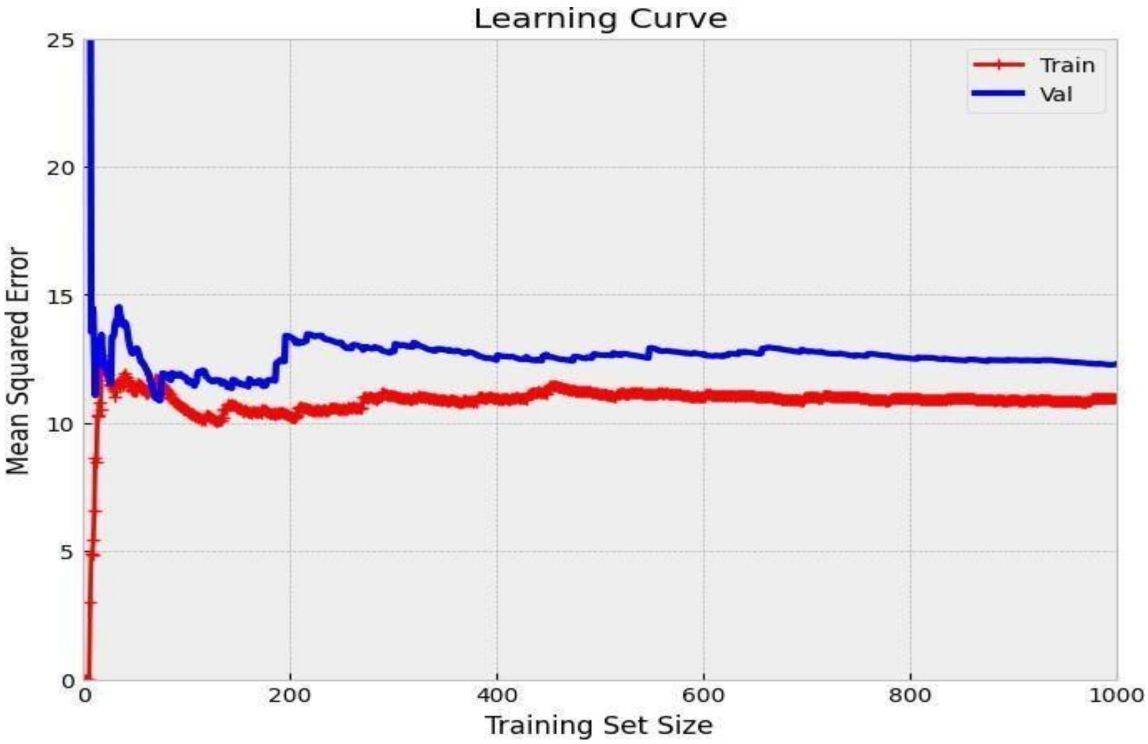
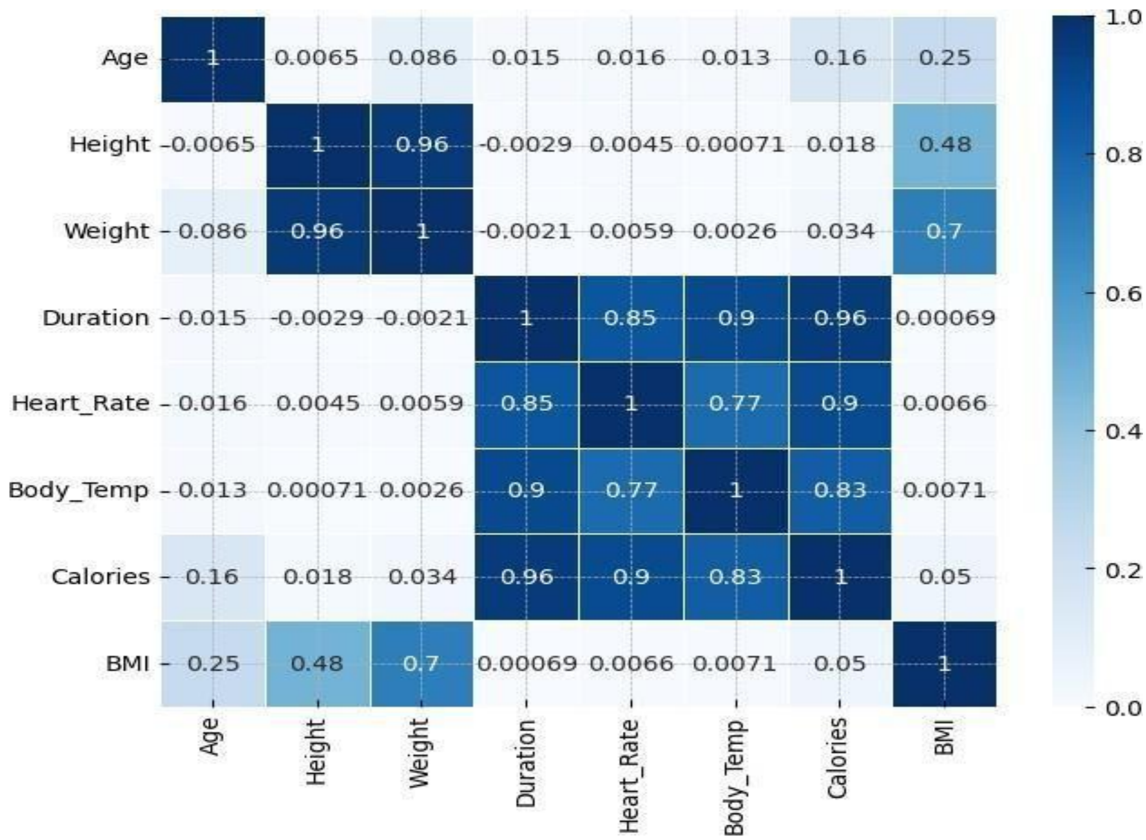
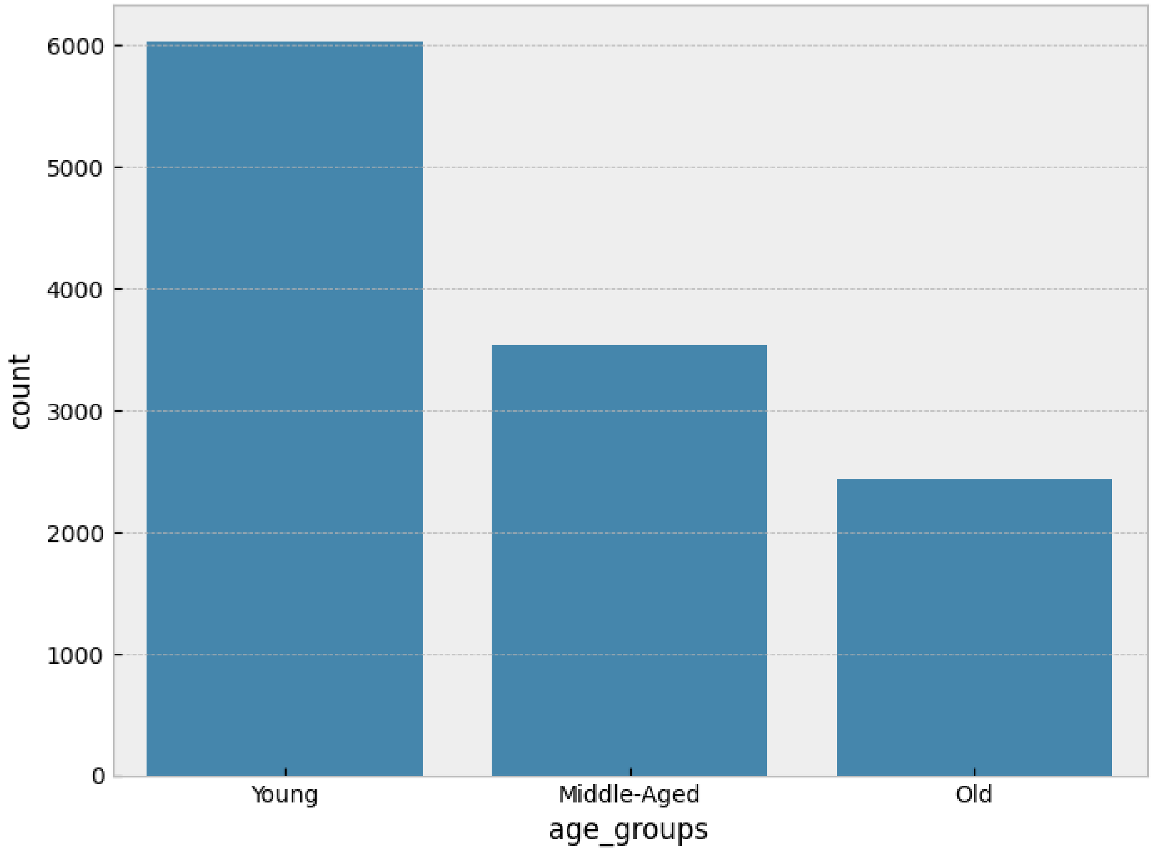


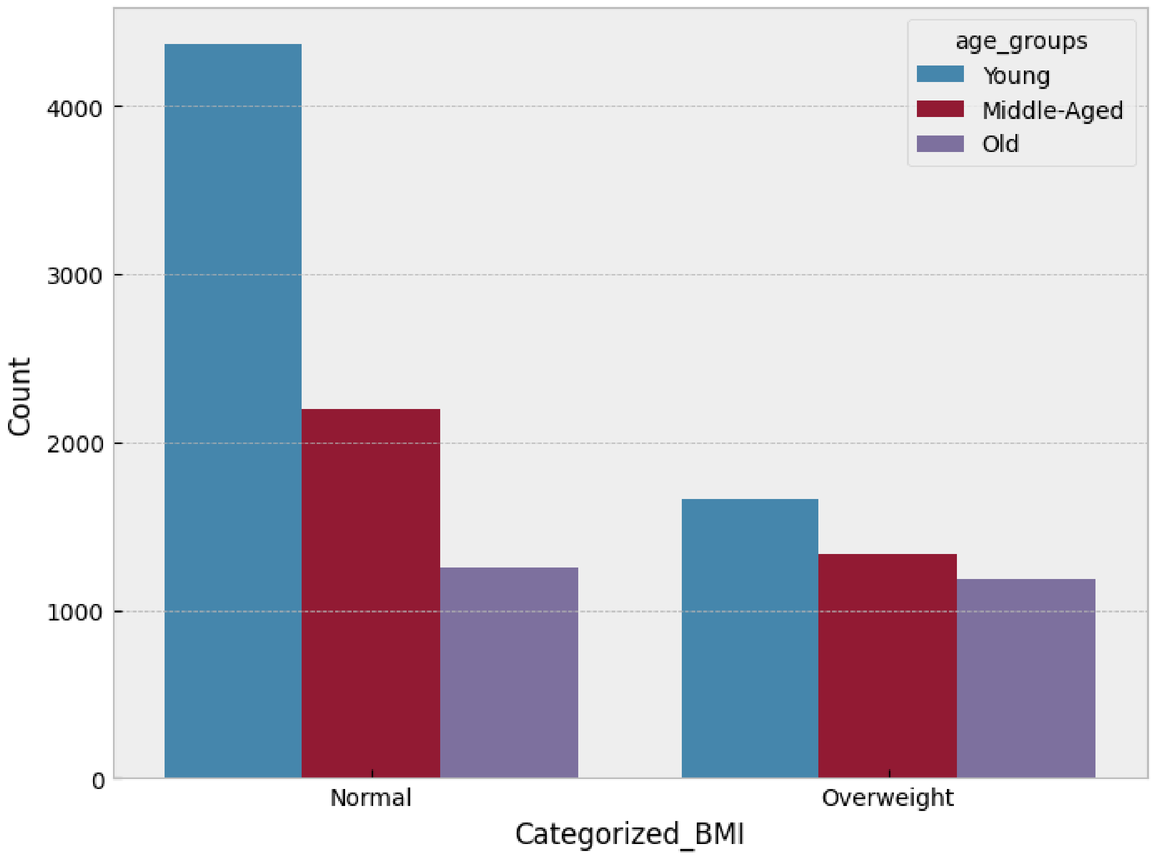
Fig 5 : Sleep Tracking & Sleep Quality Report

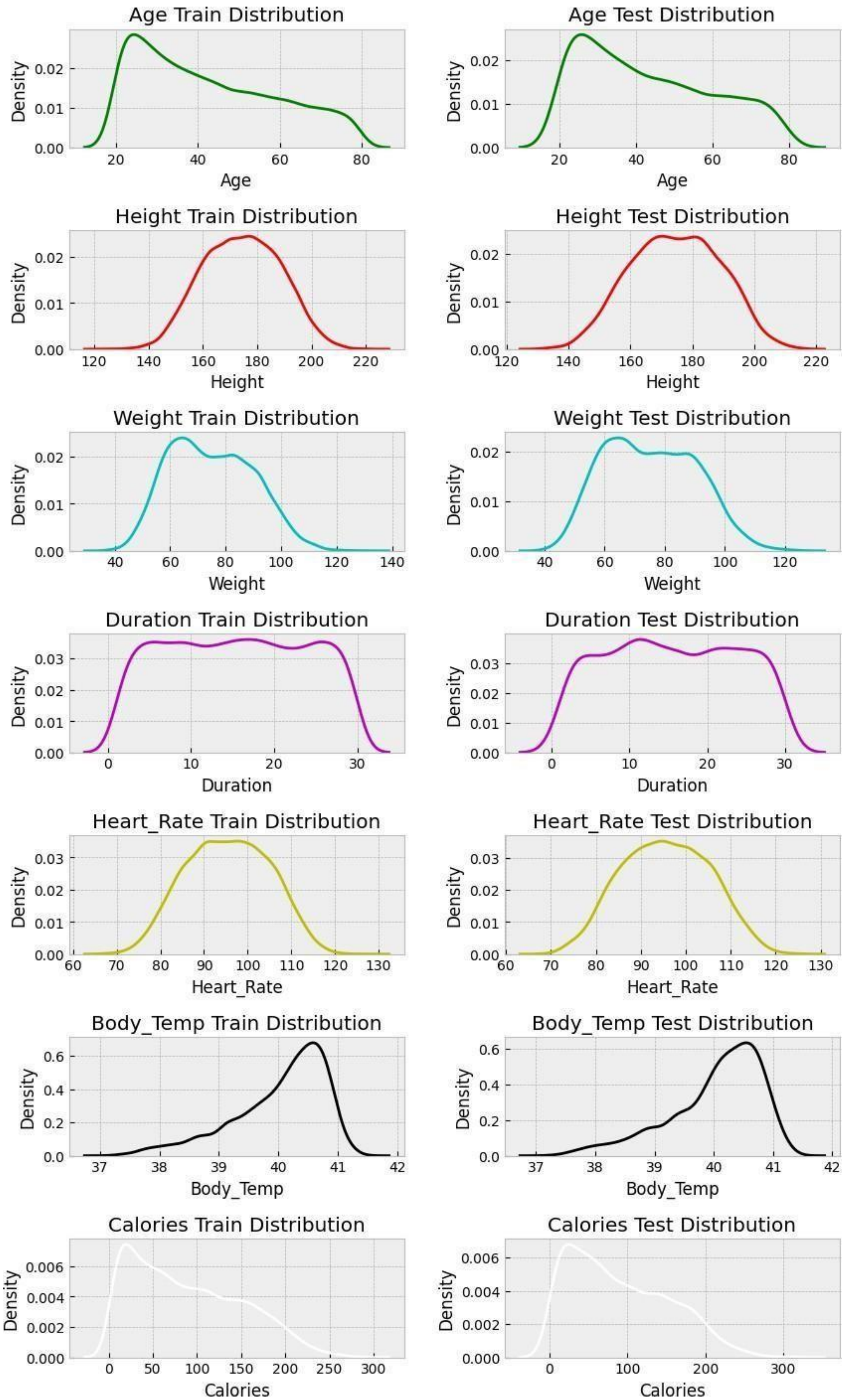
Fig 6 Exercise Performance & Workout summer

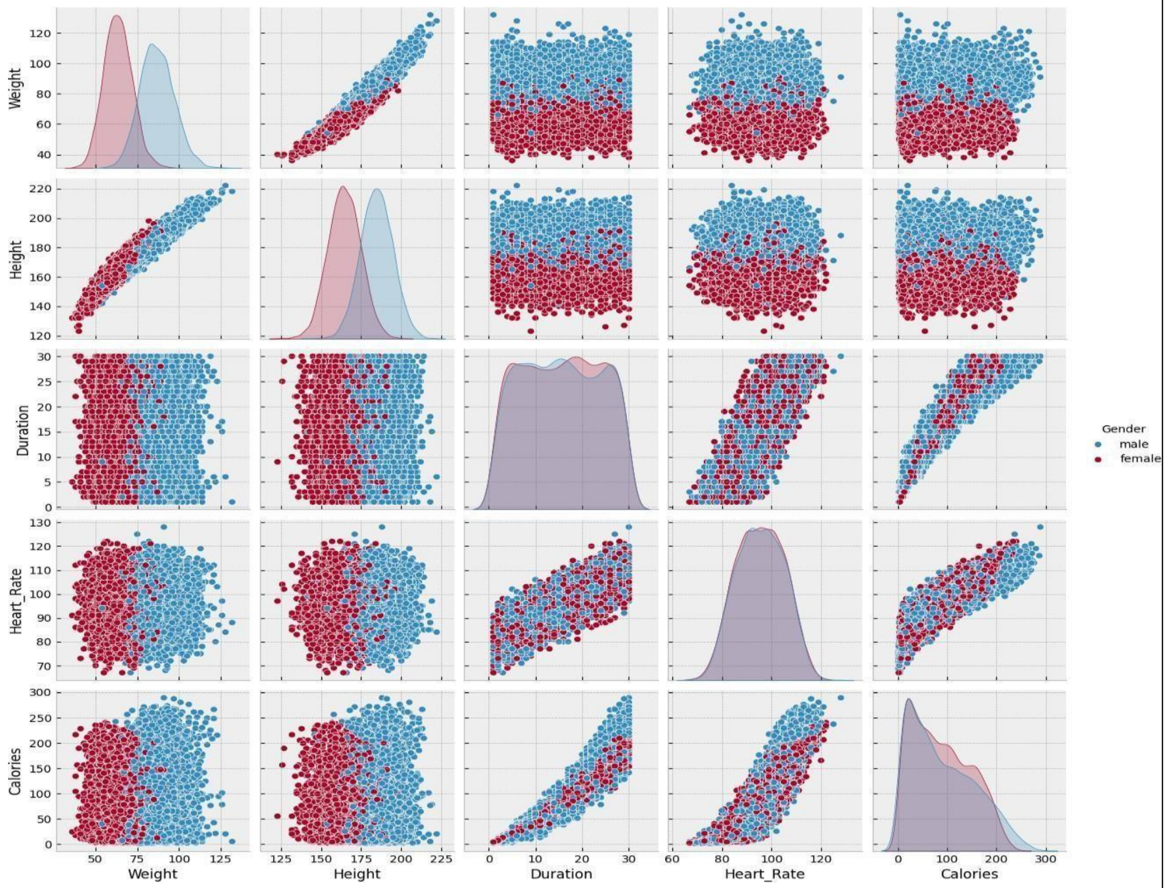


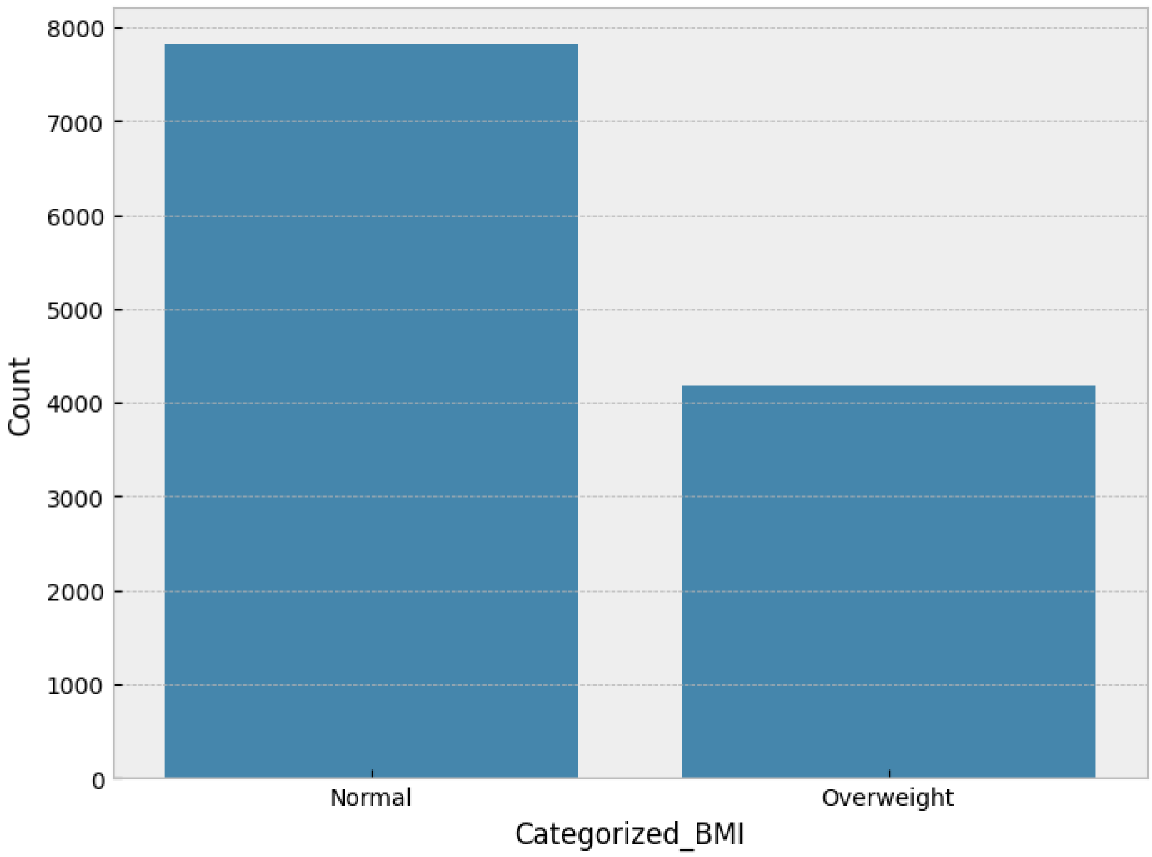



6000
5000
4000
3000
2000
1000
0
count
Young
Middle-Aged
Old
age_groups


4000
3000
2000
1000
0
Count
age_groups
Young
Middle-Aged
Old
Normal
Overweight
Categorized_BMI


Age Train Distribution
Age Test Distribution
Density
Age
Height Train Distribution
Height Test Distribution
Height
Weight Train Distribution
Weight Test Distribution
Weight
Duration Train Distribution
Duration Test Distribution
Duration
Heart_Rate Train Distribution
Heart_Rate Test Distribution
Heart_Rate
Body_Temp Train Distribution
Body_Temp Test Distribution
Body_Temp
Calories Train Distribution
Calories Test Distribution
Calories


Weight
Height
Duration
Heart_Rate
Calories
Gender
male
female


8000
7000
6000
5000
4000
3000
2000
1000
0
Count
Normal
Overweight
Categorized_BMI

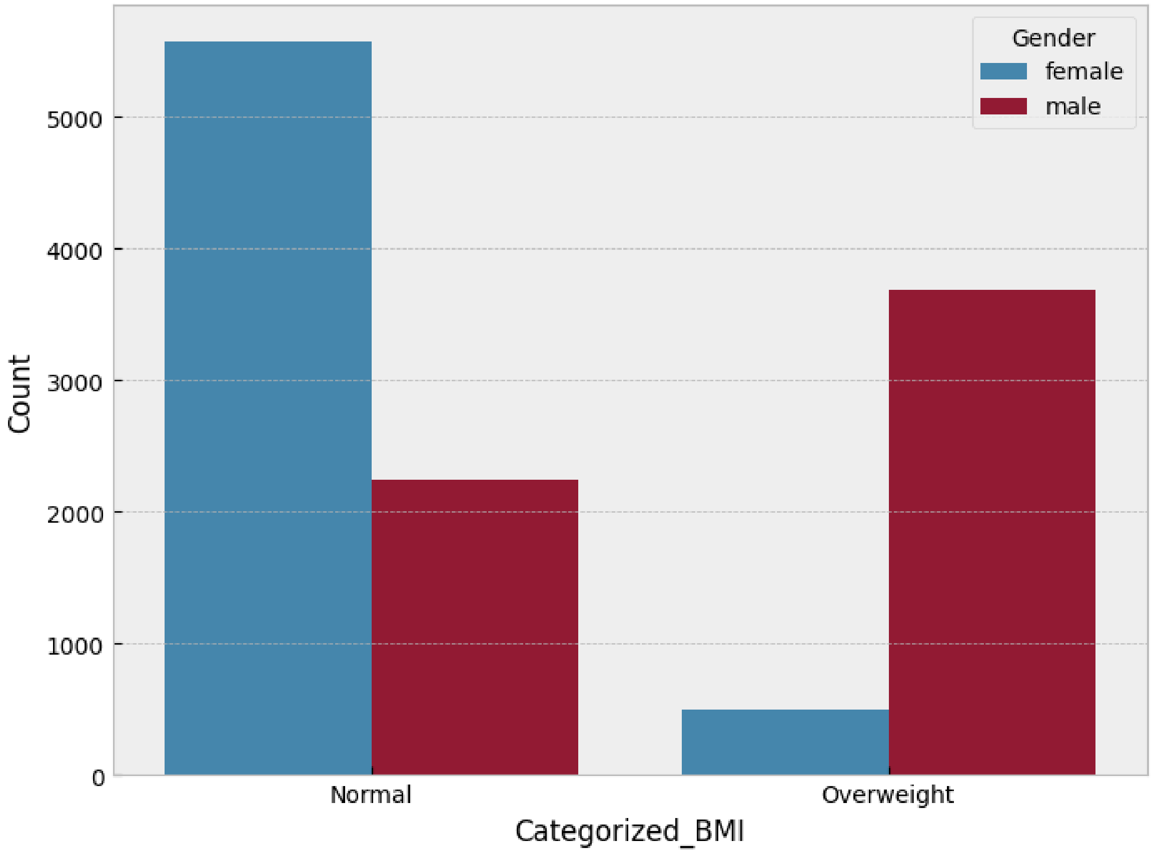
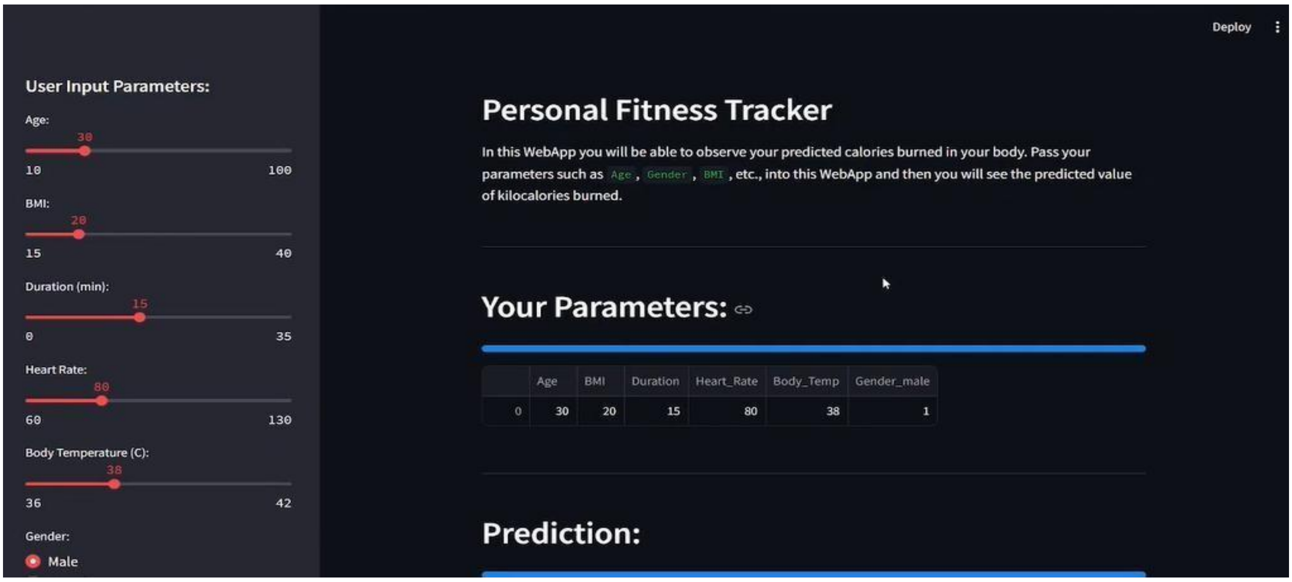


Fig 13 : Output view in app.py



4.2 GitHub Link for Code:

https://github.com/GaganB982006Hello/Gagan.B

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

While the current model provides valuable insights into fitness tracking data, several areas can be explored to improve its accuracy and effectiveness:

1. **Enhanced Data Collection** – Incorporating additional biometric data such as heart rate variability, sleep patterns, and hydration levels can improve the model's predictive capabilities.
2. **Personalized Fitness Recommendations** – Future models could integrate machine learning techniques to offer personalized fitness and diet recommendations based on users' historical data and activity trends.
3. **Inclusion of More Demographics –** Expanding the dataset to include a more diverse range of age groups, genders, and lifestyles would enhance the generalizability of the results.
4. **Real-Time Monitoring and Feedback** – Implementing a real-time analysis system to provide instant feedback and alerts for users regarding their fitness progress and potential health risks.
5. **Improving Model Accuracy** – Exploring advanced machine learning models such as deep learning or hybrid models can help improve classification accuracy and prediction reliability.
6. **Addressing Data Imbalance** – If certain groups (e.g., older individuals) are underrepresented in the dataset, future work should aim to collect more balanced data to avoid biases in model predictions.
7. **Integration with Wearable Devices** – Enhancing compatibility with multiple wearable devices and mobile applications can improve data collection and user engagement.
8. **Longitudinal Analysis** – Conducting long-term studies to analyze trends over time and understand how fitness behaviors evolve with age and lifestyle changes.

By addressing these aspects, future research can enhance the effectiveness of fitness tracking models and provide users with more accurate, actionable health insights.

5.2 Conclusion:

This project successfully analyzed fitness tracker data to gain valuable insights into users' health and activity patterns. By categorizing BMI, evaluating age-group distributions, and assessing activity levels, the model provided a comprehensive understanding of fitness trends across different demographics.

The study highlighted key differences in fitness behaviors among various age groups, offering insights into how lifestyle choices impact health. The integration of fitness tracking data with analytical models demonstrated the potential for data-driven decision-making in personal health management.

Furthermore, the project contributes to the growing field of wearable technology research by emphasizing the importance of continuous monitoring and personalized fitness recommendations. The findings can aid individuals in making informed lifestyle choices while also assisting health professionals in designing targeted wellness programs.

The project emphasizes the potential for integrating machine learning techniques to further refine health predictions and activity recommendations. By leveraging real-time fitness data, future advancements can provide more accurate assessments of users' health *risks and personalized workout plans.*

Moreover, this study opens the door for collaboration between wearable technology developers, healthcare professionals, and data scientists to enhance user experience and promote healthier lifestyles. As fitness tracking devices continue to evolve, incorporating AI-driven insights and behavioral analytics will pave the way for smarter, more adaptive health monitoring systems.

REFERENCES

[1]. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002

[2]. on Streamlit Cloud, 2024.Now you have a total of seven references formatted as per your request. Let me know if you need any modifications or additi[1]. AICTE Internship 2024 Project Report, "Fitness Tracker using AI," TechSaksham – A joint CSR initiative of Microsoft & SAP.

[3]. app.py – Python-based implementation of a fitness tracker, utilizing machine learning for real-time calorie prediction, 2024.

[4]. fitness_tracker.ipynb – Data processing and visualization notebook, analyzing exercise trends and fitness metrics, 2024.

[5]. Faulkner Act (Mayor-Council System), New Jersey Government Structure Documentation, 2024.

[6]. Random Forest Regressor Model, "Implementation of Machine Learning for Calorie Prediction," Developed using Scikit-Learn, 2024.

[7]. Streamlit Web Application, "Interactive User Interface for Fitness Tracking," Deployment onal references!