**AI Powered Personal Fitness Tracker**

A Project Report submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning

with

Tech Saksham – A joint CSR initiative of Microsoft & SAP

By

**Adibhatla Vaishnavi Amulya**

**vaishnaviamulyaadibhatla@gmail.com**

Under the Guidance of

**Saomya Chaudhury**

**ACKNOWLEDGEMENT**

I would like to use this moment to thank all those who have guided and supported me during the duration of this internship project, directly and indirectly, for their valuable input. Above all, I am deeply grateful to my supervisor, **Mr. Saomya Chaudhury**, for his outstanding guidance, unwavering support and, constructive feedback. His mentorship and encouragement has been an integral part of the successful completion of this project. I would also like to thank **Tech Saksham, a collaborative CSR initiative of SAP and Microsoft**, for giving me this incredible learning experience through the **AICTE Internship on AI: Transformative Learning.** The initiative provided a great platform to learn through practical experience and increase my knowledge in the field of artificial intelligence. Lastly, I extend my gratitude to all the faculty members, mentors and well-wishers who have contributed in any way to my successful completion of this project.

**Adibhatla Vaishnavi Amulya**

**ABSTRACT**

This project titled **AI Powered Personal Fitness Tracker** aimed at examining exercise and health data to forecast the fitness age of a person based on numerous machine learning models. The research incorporates datasets that have user exercise information and calorie intake to investigate the primary health parameters such as heart rate, BMI, body temperature, and calorie burn rate. The process includes data preprocessing with dataset merging, missing values handling, feature engineering (calculating metrics such as Heart Rate Recovery Index and Calories Burn Rate), and visualization using scatter plots, pie charts, boxplots, and heatmaps, e.t.c. The dataset is also used in predictive modeling where features such as gender, height, weight, duration, and heart rate are used as input to train regression models. A variety of machine learning algorithms such as **Linear Regression, Ridge, Lasso, Decision Tree, Random Forest, Support Vector Regression, and Gradient Boosting** were compared on the basis of performance metrics such as **Mean Absolute Error (MAE), Mean Squared Error (MSE), and R² score.** The Gradient Boosting Regressor was the highest-scoring model with an R² value of 77.97%, signifying a very good predictive power. The project has also been provided with an interactive Streamlit web app that will enable users to provide their information and receive a fitness age estimate. In summary, the project is able to convincingly illustrate the use of data-driven intelligence and machine learning in facilitating the monitoring and enhancement of one's fitness. Real-time tracking and personalized advice on fitness would be useful additions for enhanced health management in the future.

**TABLE OF CONTENTS**

**Abstract 3**

**Chapter 1.**  **Introduction 7**

1.1 Problem Statement 7

1.2 Motivation 7

1.3 Objectives 8

1.4. Scope of the Project 8

**Chapter 2.**  **Literature Survey 10**

2.1 Introduction 10

2.2 Review of Existing Models 10

2.3 Gaps and Limitations in Existing Models 11

2.4. Proposed Solution and Justification 12

**Chapter 3.**  **Proposed Methodology 13**

3.1 System Design 13

3.2 Requirement Specifications 16

**Chapter 4.**  **Implementation and Results**  **18**

4.1 Snap Shot of Results 18

4.2 GitHub Link for Code 21

**Chapter 5. Discussion and Conclusion 22**

5.1 Future Work 22

5.2 Conclusion 22

**References** 24

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | System Design | 13 |
|  | Data preprocessing | 14 |
|  | Pair Plot | 18 |
|  | Correlation Plot | 19 |
|  | App Output - 01 | 20 |
|  | App Output - 02 | 20 |
|  | App Output - 03 | 21 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table. No.** | **Table Caption** | **Page No.** |
| **3.1** | DataSet | 14 |
| **3.2** | ML Model Performance | 16 |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

Maintaining a healthy lifestyle requires a balanced approach to exercise, nutrition, and the monitoring of vital health metrics. However, many individuals lack the tools to effectively track their progress and gain a comprehensive understanding of their overall health. Traditional fitness tracking methods often provide surface-level data, such as step counts or calories burned, without offering deeper insights into an individual's true fitness level. One key indicator of overall health is fitness age, which estimates the body’s biological age based on health and activity data rather than chronological age. Without leveraging advanced analytics and machine learning, individuals face challenges in accurately assessing their fitness levels and optimizing their routines. This project aims to tackle this challenge by developing an AI-powered fitness tracker that uses health parameters, exercise patterns, and dietary habits to predict an individual’s fitness age. By offering data-driven insights, the system empowers users to make informed decisions, improve their fitness, and lead healthier lives

* 1. **Motivation:**

The inspiration for this project comes from the increasing concern for the necessity of living a healthy lifestyle in conjunction with physical exercise, proper nutrition, and general well-being. Even as there has been an explosion in fitness tracking technologies, most available solutions offer just superficial information in the form of steps taken or calories burned without providing meaningful insight into a person's overall state of health. The fitness age, a measure that indicates the biological health of the body as opposed to its age, provides a more significant measurement to determine overall fitness. Yet with the lack of smart systems utilizing AI and machine learning, people are unable to make an accurate assessment of their own fitness levels or customize their routines accordingly. This project aims to fill this gap by designing an AI-driven fitness tracker. The system enables users by applying sophisticated analytics to forecast fitness age, allowing them to better comprehend their health today, set realistic targets, and maximize workout and nutrition routines. The motivation also comes from cultivating a proactive response to health and motivating people to make smart choices that can drive long-term well-being and a healthier tomorrow. By fulfilling this unmet need, the project has the potential to have a major impact, assisting individuals in attaining personal fitness objectives while inspiring a wholesome appreciation of health.

* 1. **Objective:**

The primary objectives of this project are:

* To create an AI-driven personal fitness tracker offering a better insight into overall fitness and health.
* To estimate a person's fitness age, a health measure based on health markers, exercise habits, and eating habits, as opposed to chronological age.
* To use sophisticated data analytics and machine learning algorithms to create insightful information for users.
* To equip users with actionable, data-driven insights to inform their fitness and lifestyle decisions.
* To combine several metrics, including BMI, heart rate, and calorie burn, to provide a well-rounded and personalized fitness perspective.
* -To promote healthier habits and inspire users through individualized fitness objectives.
  1. **Scope of the Project:**

The scope of this project is defined by the following:

* Develop an AI-powered fitness tracker to provide a comprehensive understanding of users' health and fitness.
* Predict fitness age based on key health parameters, including BMI, heart rate, calorie burn, exercise patterns, and dietary habits.
* Offer personalized insights and tailored recommendations to help users monitor their progress effectively.
* Utilize machine learning algorithms to generate accurate, data-driven predictions and feedback.
* Provide applications for individuals to manage personal health and support fitness trainers or healthcare professionals in designing customized plans.
* Promote long-term fitness and healthier lifestyles through goal setting, progress tracking, and motivational insights.
* Address limitations such as dependence on user-provided data and external factors while focusing on potential future enhancements.

**CHAPTER 2**

**Literature Survey**

1. **Introduction:**

Fitness tracking and health monitoring have become increasingly important in recent years. With advancements in data science and machine learning, it is now possible to predict an individual’s fitness age based on various health indicators such as heart rate, body mass index (BMI), and exercise duration. This project aims to develop a Personal Fitness Tracker that utilizes machine learning models to predict fitness age based on physiological and exercise-related parameters. The application provides valuable insights into an individual's overall health and fitness level, helping users make informed lifestyle decisions.

1. **Review of Existing Models:**

In this project, multiple regression-based models have been explored for predicting fitness age. The following machine learning models were implemented and evaluated:

1. **Linear Regression:** A simple, interpretable model that assumes a linear relationship between input features and the target variable.
2. **Ridge and Lasso Regression**: Regularized linear regression models that help in reducing overfitting by adding penalty terms to the cost function.
3. **ElasticNet Regression:** A combination of Ridge and Lasso that balances regularization and feature selection.
4. **Random Forest Regressor:** An ensemble learning method that uses multiple decision trees to improve predictive performance and reduce variance.
5. **Gradient Boosting Regressor:** A boosting algorithm that improves weak learners iteratively to enhance accuracy.
6. **AdaBoost Regressor:** Another boosting method that assigns higher weights to poorly predicted instances to improve performance.
7. **Support Vector Regressor (SVR):** A regression model that finds the best hyperplane to minimize prediction error using support vectors.
8. **K-Nearest Neighbours (KNN) :** A non-parametric model that predicts values based on the closest training data points.
9. **Decision Tree Regressor:** A tree-based model that recursively splits data based on the best possible feature.

The models were evaluated based on Mean Absolute Error (MAE), Mean Squared Error (MSE), and R² Score.

1. **Gaps and Limitations in Existing Models:**

While the explored models provide reasonable accuracy, they have inherent limitations:

1. **Linear Regression, Ridge, and Lasso Regression:**
   * These models assume a linear relationship, which may not fully capture the complexities of fitness age prediction.
   * Poor performance in handling non-linear relationships between input features and target variables.
2. **Decision Tree and Random Forest Regressors:**
   * Decision Trees tend to overfit the training data.
   * Random Forest, while reducing overfitting, still struggles with complex dependencies.
3. **KNN and SVR:**
   * KNN requires a large amount of memory and is computationally expensive for large datasets.
   * SVR is highly sensitive to hyperparameters and requires careful tuning.
4. **Boosting Models (Gradient Boosting, AdaBoost):**
   * These models improve prediction accuracy but can be computationally expensive.
   * AdaBoost performed poorly due to its sensitivity to noisy data.
5. **Proposed Solution and Justification:**

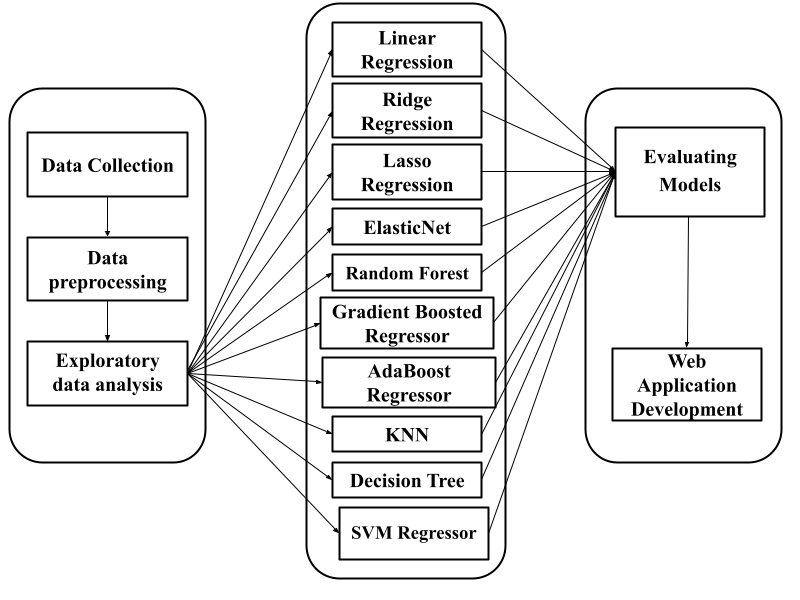
Based on the analysis, the Gradient Boosting Regressor was selected as the best-performing model for predicting fitness age. The justification for this choice includes:

1. **High Predictive Performance:**
   * Achieved the highest R² score (77.97%) compared to other models.
   * Effectively captures complex non-linear relationships between input features.
2. **Robustness to Outliers:**
   * Unlike Linear Regression, Gradient Boosting is less sensitive to outliers in the dataset.
3. **Feature Importance Analysis:**
   * Gradient Boosting allows us to determine the most important features contributing to fitness age prediction, improving model interpretability.
4. **Generalization Capabilities:**
   * Unlike Decision Trees, it does not overfit easily, making it more reliable for unseen data.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design:**

****

**Figure 1: System Design**

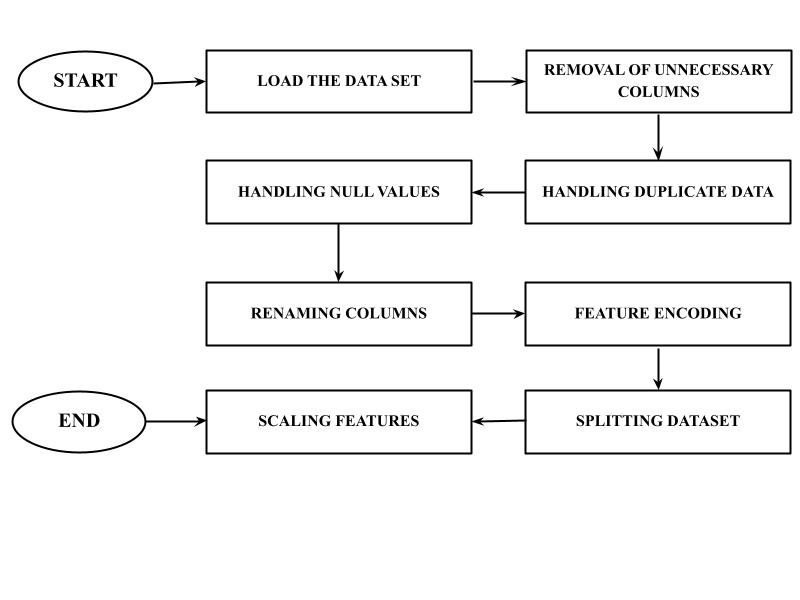
1. **Data Collection:**

The dataset Initially contained 15,000 rows and 9 columns. Through a comprehensive cleaning and preprocessing process, the dataset was refined to 14999 rows and 12 columns. Key attributes include Age, Height, Weight, BMI, Heart rate etc. all of which are crucial for analyzing factors.

|  |  |  |
| --- | --- | --- |
| **S.NO** | **COLUMN NAME** | **DATA TYPE** |
| **1** | User\_ID | int64 |
| **2** | Calories | float |
| **3** | Gender | object |
| **4** | Age | int64 |
| **5** | Height | float |
| **6** | Weight | float |
| **7** | Duration | float |
| **8** | Heart\_Rate | float |
| **9** | Body\_Temp | float |

**Table 3.1: DataSet**

1. **Data Preprocessing:**



**Figure 2: Data Preprocessing**

To ensure the dataset's quality and relevance for analysis, several data preprocessing steps were undertaken. The data preprocessing phase began with importing essential Python libraries, including NumPy, Pandas, Plotly, Seaborn, Matplotlib, and Scikit-learn. The datasets exercise.csv and calories.csv were loaded and merged on the common column User\_ID to create a unified dataset. Duplicate and missing values were checked, and the User\_ID column was removed. Column names were also standardized for better readability. Feature engineering was performed by deriving new attributes such as Body Mass Index (BMI) using weight and height, Heart Rate Recovery Index calculated as the ratio of heart rate to age, and Calories Burn Rate, which was obtained by dividing calories burned by exercise duration.

1. **Exploratory Data Analysis:**

The Exploratory Data Analysis (EDA) phase involved analyzing data distributions, identifying patterns, and detecting outliers. Summary statistics and visualizations such as bar charts, pie charts, scatter plots, and box plots were used to explore variables like age, gender, exercise duration, calories burned, and body temperature. A correlation heatmap and pair plots helped identify relationships between numerical features. These analyses helped in understanding trends and preparing the data for modeling.

1. **Machine Learning Model Training:**

The dataset was split into training and testing sets to build predictive models. Various machine learning algorithms, including linear regression, decision trees, and random forests, were implemented to predict calorie burn based on factors such as exercise duration and body temperature. Feature scaling and encoding techniques were applied where necessary to optimize model performance. Hyperparameter tuning was conducted to enhance model accuracy and prevent overfitting.

1. **Model Evaluation:**

The trained models were evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R²) scores. Cross-validation techniques were applied to ensure model reliability. The best-performing model was selected based on the lowest error values and highest predictive accuracy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.NO** | **MODEL** | **MAE** | **MSE** | **R2\_Score** |
| **1** | Gradient  Boosting Regressor | 5.87 | 61.32 | 77.97 |
| **2** | Support  Vector Regressor | 9.32 | 144.63 | 48.05 |
| **3** | K-Nearest Neighbors | 10.06 | 164.20 | 41.02 |
| **4** | Linear Regression | 10.35 | 164.48 | 40.92 |
| **5** | Ridge Regression | 10.35 | 164.49 | 40.92 |
| **6** | Lasso Regression | 10.38 | 164.82 | 40.80 |
| **7** | ElasticNet | 10.46 | 166.95 | 40.03 |
| **8** | Random  Forest Regressor | 12.04 | 211.72 | 23.95 |
| **9** | Decision  Tree Regressor | 13.10 | 246.47 | 11.47 |
| **10** | AdaBoost Regressor | 13.87 | 258.71 | 7.07 |

**Table 3.2: ML Model Performance**

1. **Web Application Development:**

A web application was developed using Streamlit to provide an interactive interface for predicting calorie burn and fitness age based on user inputs. The trained machine learning model was integrated into the application, allowing real-time predictions. The interface was designed to be user-friendly, enabling users to enter parameters such as exercise duration and body temperature. Streamlit's built-in visualization features were used to display model predictions and insights effectively. The application was deployed locally, making it accessible for testing and further improvements.

* 1. **Requirement Specifications:**

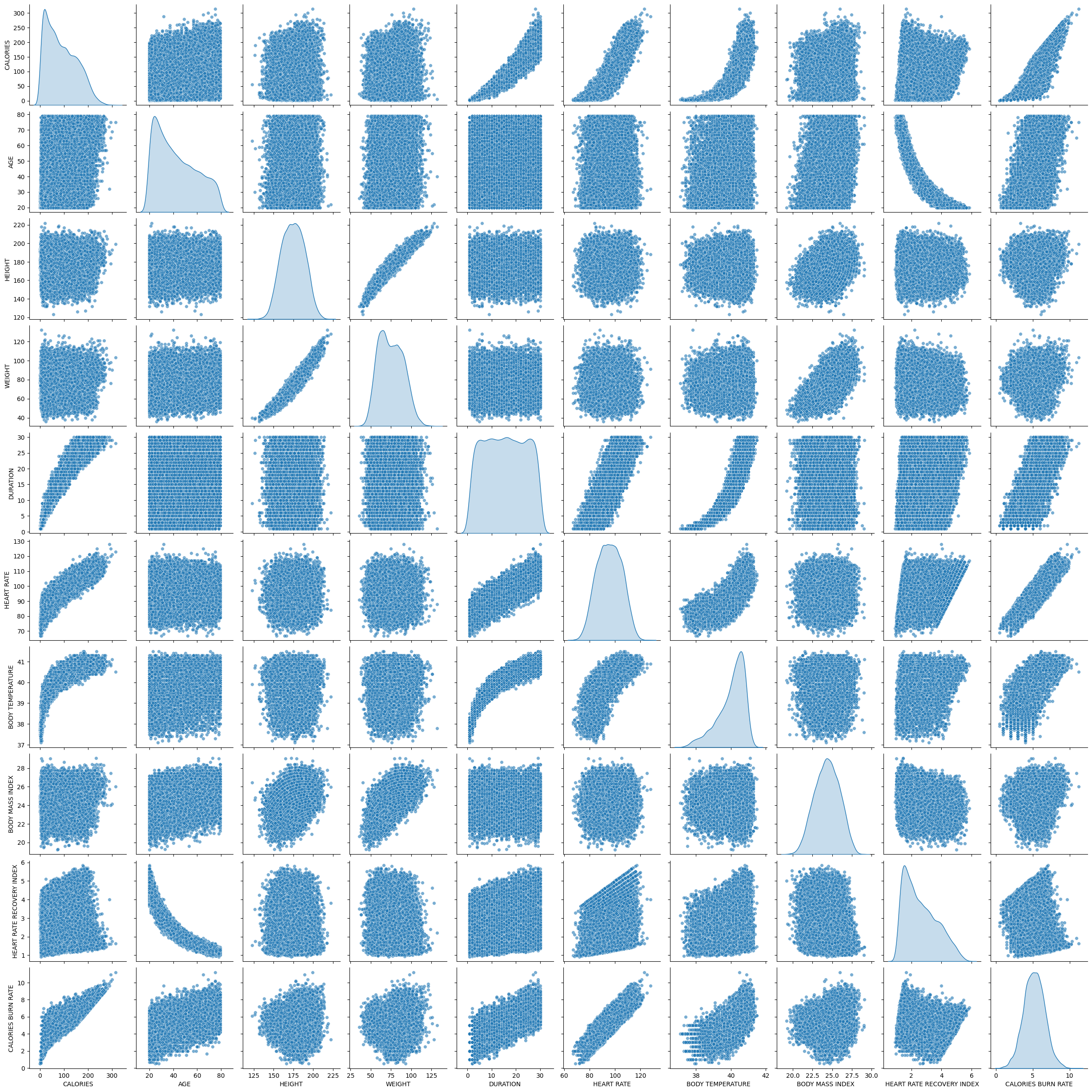
Mention the tools and technologies required to implement the solution.

* + 1. **Hardware Requirements:**
* Minimum 8 GB RAM for smooth execution.
* Multi-core CPU for parallel processing.
* Disk space for data storage and application deployment.
  + 1. **Software Requirements:**
* Python 3.x: Core language for application development.
* **Libraries:**
  + streamlit – For building the web-based user interface.
  + pandas – For data manipulation and preprocessing.
  + numpy – For numerical operations.
  + scikit-learn – For model development and evaluation.
  + matplotlib and seaborn – For data visualization.
  + IDE: Jupyter Notebook for model development and initial testing.
  + Deployment Environment: Localhost or cloud server for application hosting.

**CHAPTER 4**

**Implementation and Result**

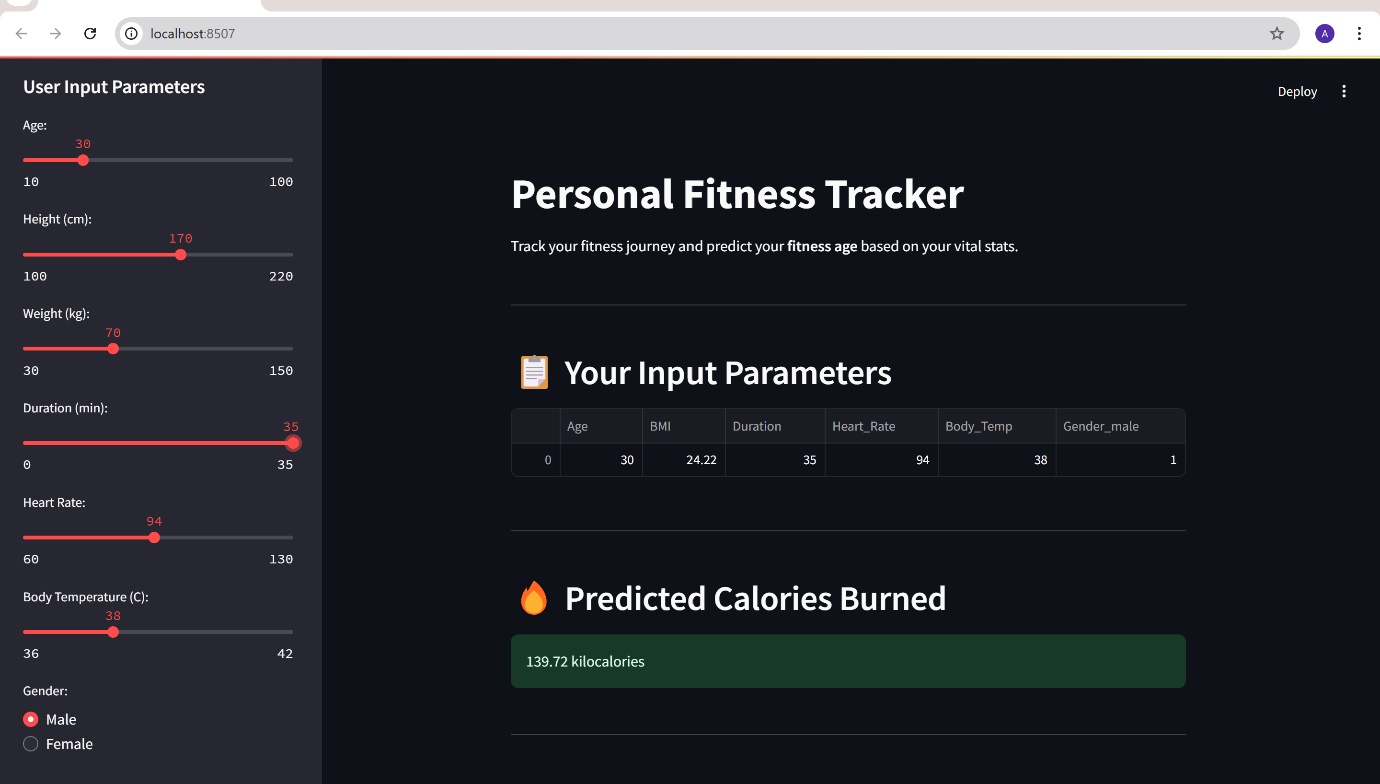
* 1. **Snap Shots of Results:**

****

**Figure 3: Pair Plot**

****

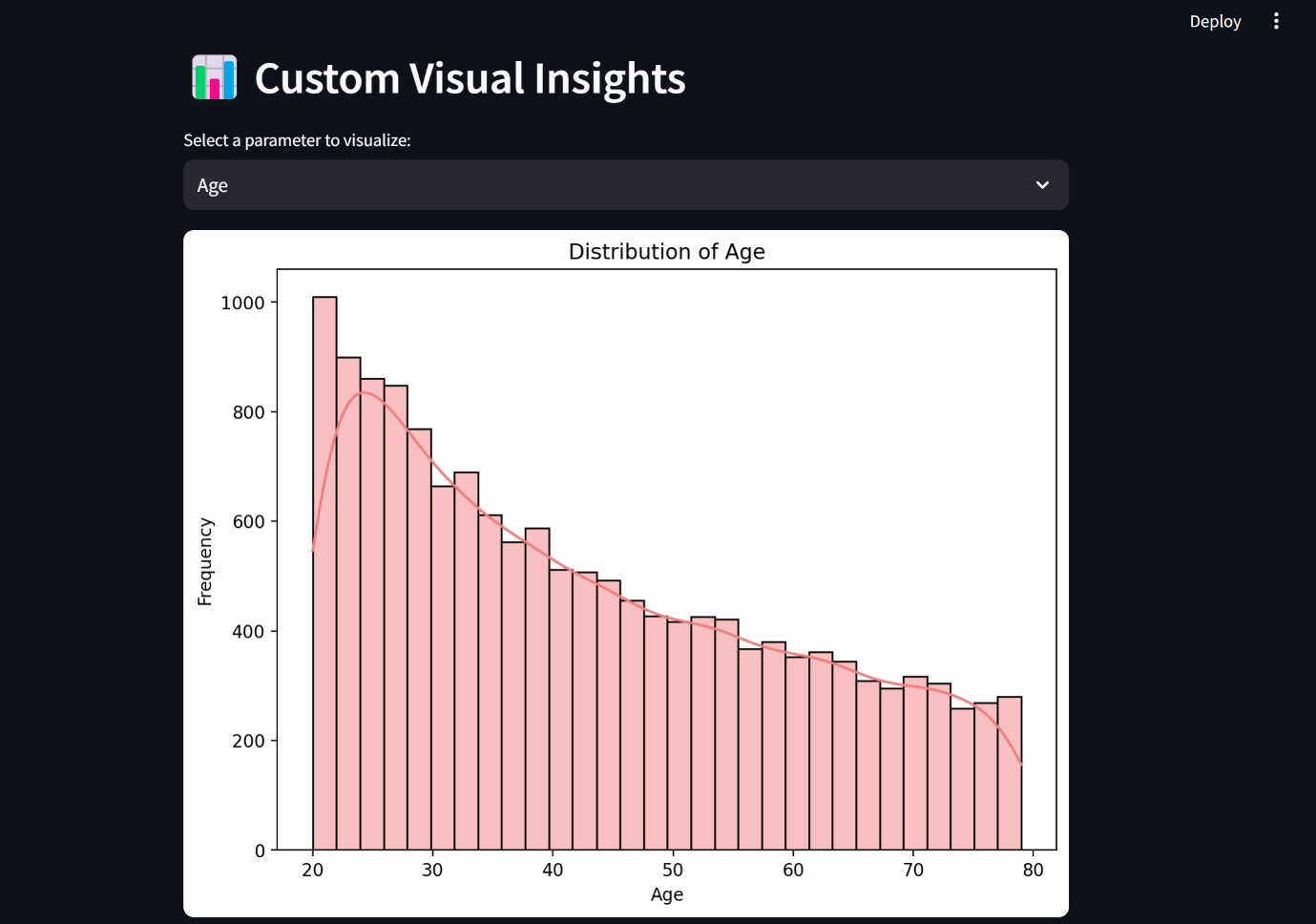
**Figure 4: Corelation Plot**

****

**Figure 5: App Output - 01**

****

**Figure 6: App Output - 02**

****

**Figure 7: App Output - 03**

* 1. **GitHub Link for Code:**

<https://github.com/Adibhatla-Vaishnavi-Amulya/Personal_Fitness_Tracker.git>

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

The future scope of this project includes several enhancements to improve the model's performance and address existing limitations:

1. **Expand Dataset:** Incorporate a larger and more diverse dataset to improve the model's generalization capabilities across varied demographics and fitness levels.
2. **Include Additional Parameters:** Integrate more health metrics, such as VO2 max, sleep quality, hydration levels, and stress indicators, to enhance prediction accuracy.
3. **Optimize Computational Efficiency:** Focus on reducing the computational complexity of Gradient Boosting for real-time predictions and scalability.
4. **Develop Personalized Recommendations:** Build a recommendation engine that suggests tailored workout routines, dietary plans, and lifestyle changes based on fitness age predictions.
5. **Real-Time Tracking:** Incorporate wearable device integration to enable continuous data collection and dynamic updates of fitness age predictions.
6. **Interactive Visualizations:** Enhance the system interface with intuitive and interactive charts to provide deeper insights into user progress and health trends.
7. **Address Model Bias:** Evaluate and mitigate potential biases in the model, ensuring fairness across diverse populations.
8. **Mobile Application:** Extend the fitness tracker as a mobile app to increase accessibility and user engagement.
   1. **Conclusion:**

This project makes significant contributions toward bridging the gap between traditional fitness tracking tools and advanced health monitoring systems. By utilizing machine learning, the system predicts fitness age, providing users with a more meaningful and personalized metric to understand their biological health. The integration of health indicators, exercise patterns, and dietary habits empowers individuals to make informed decisions about their lifestyle and fitness goals. The adoption of the Gradient Boosting Regressor ensures robust predictive performance, addressing complex non-linear relationships in the data. While the current implementation achieves strong accuracy, the proposed future work highlights opportunities for improvement, scalability, and enhanced user experience. Ultimately, the project fosters a proactive approach to health management and promotes long-term well-being, paving the way for innovative applications in personal fitness and healthcare domains.

**REFERENCES**

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.
2. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 12, 2825–2830.
3. Friedman, J. H. (2001). *Greedy Function Approximation: A Gradient Boosting Machine.* Annals of Statistics, 29(5), 1189-1232.
4. Hunter, J. D. (2007). *Matplotlib: A 2D Graphics Environment.* Computing in Science & Engineering, 9(3), 90-95.
5. Waskom, M. L. (2021). *Seaborn: Statistical Data Visualization.* Journal of Open Source Software, 6(60), 3021.
6. McKinney, W. (2010). *Data Structures for Statistical Computing in Python.* Proceedings of the 9th Python in Science Conference, 51-56.
7. Streamlit Documentation. (2024). *Build Web Applications Using Streamlit.* Retrieved from https://docs.streamlit.io/