Dr. AMBEDKAR INSTITUTE OF TECHNOLOGY

Near Jnana Bharathi Campus, Bengaluru-560 056

(An Autonomous Institution, Aided by Government of Karnataka)

Project Report

on

"CALORIE BURNT PREDICTION"

Submitted by:

MUKTHI H K	NISARGA K	PALLAVI S D
[1DA21CS091]	[1DA21CS093]	[1DA21CS098]

Under the guidance Of

Arathi P	Dr . Madhu B
Asst. Prof., Dept of CSE	Asst. Prof., Dept of CSE
Dr. AIT	Dr. AIT

Department of Computer Science and Engineering

2023-24

Dr. AMBEDKAR INSTITUTE OF TECHNOLOGY

Near Jnana Bharathi Campus, Bengaluru-560 056

(An Autonomous Institution, Aided by Government of Karnataka)

CERTIFICATE

This is to certify that the project entitled **"Calorie burnt prediction"** submitted in the partial fulfillment of the requirement of the 6th semester AIML project curriculum during the year 2023- 2024 is a result of bonafide work carried out by

MUKTHI H K	**NISARGA K**	**PALLAVI S D**
[1DA21CS091]	**[1DA21CS093]**	**[1DA21CS098]**

Signature of the guide:

_________________ _________________

Arathi P	**Dr . Madhu B**
Asst. Prof., Dept of CSE	**Asst. Prof., Dept of CSE**
Dr. AIT	**Dr. AIT**

1. External Examiner ________________
2. Internal Examiner____

Dr. Siddaraju
H.O.D
Dept of CSE, Dr.AIT

ACKNOWLEDGEMENT

The satisfaction that accompanies this project would be incomplete without the mention of the people who made it possible, without whose constant guidance and encouragement would have made our efforts go in vain.

We consider ourselves privileged to express our gratitude and respect toward all those who guided us through the project, "**Calorie burnt prediction**".

We would like to express our gratitude to **C.Nanjundaswamy, Principal, Dr. AIT**, for providing us the congenial environment to work in.

We would like to express profuse gratitude to **Dr. SIDDARAJU, HOD, Dept. of Computer Science & Engineering, Dr.AIT**, for giving us the support, encouragement and providing us the required lab facilities that was necessary for the completion of this project.

As a token of gratitude, we would like to acknowledge our sincere gratefulness to the internal guides

Dr. Madhu B, Asst. Professor, Dept. of CSE, DR. AIT and Arathi P, Asst. Professor, Dept. of CSE, DR. AIT for their unlimited support and encouragement and environment provided throughout the process.

We also express our gratitude and sincere thanks to all the teaching and non-teaching staff of Computer Science & Engineering Department.

Finally, yet importantly, we would like to express our heartfelt thanks to our **Parents** for their blessings and our **Friends** for their help and wishes for the successful completion of this project report.

ABSTRACT

Calorie burnt prediction project aims to develop a sophisticated calorie burn prediction model utilizing advanced Artificial Intelligence (AI) and Machine Learning (ML) techniques. The goal is to accurately estimate the number of calories burned during various physical activities by leveraging a comprehensive set of input features, including demographic details (age, weight, height, gender) and physiological metrics (heart rate, body temperature and duration).

The project uses a diverse dataset capturing a wide range of individual profiles and activity patterns to train and validate the Random Regression Model. The optimized model will be integrated into a user-friendly application using Streamlit, enabling users to input their personal and activity data to receive real-time calorie burn estimates.

This application, powered by Streamlit, caters to fitness enthusiasts, healthcare providers, and individuals keen on monitoring and managing their physical activity levels effectively. The project's results demonstrate the effectiveness of AI and ML, specifically the Random Regression Model, in improving calorie burn prediction accuracy over traditional methods. This underscores the potential of these technologies in offering personalized health insights and supporting better health and fitness management.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

FIGURES:

TABLES:

CHAPTER 1

INTRODUCTION

1.1 Problem statement

In contemporary health and fitness landscapes, accurately estimating calorie expenditure during physical activity remains a significant challenge. Existing methods often rely on generalized formulas or wearable technologies, which overlook individual nuances and fail to provide precise estimations. Recognizing this gap, our project endeavors to pioneer a novel approach using advanced machine learning techniques to develop a comprehensive predictive model.

By synthesizing extensive datasets encompassing diverse individual attributes (such as age, weight, height, gender) and enhanced activity parameters (including duration, intensity and body temperature). our model aims to provide personalized and finely tuned estimations of calorie expenditure. Through this endeavor, we aspire to empower individuals to embark on their fitness journeys with confidence, armed with the knowledge needed to optimize their workouts, track progress accurately.

1.2 Scope of the project

- **Data Collection:** The project will involve gathering a diverse range of data sources containing information on individual characteristics (such as age, weight, height, gender) and activity parameters (including duration, intensity, body temperature). This data will serve as the foundation for training and validating the machine learning models.

- **Machine Learning Models:** The project scope entails exploring and implementing various machine learning algorithms using regression, and to predict calorie burn. These models will undergo rigorous training, optimization, and evaluation to ensure accuracy and generalizability. Ultimately, the project aims to develop robust and accurate models that empower individuals to make informed decisions about their fitness goals.

- **Feature Engineering:** Feature engineering techniques will be employed to extract relevant features from the dataset, including transformations, scaling, and encoding of categorical variables. This process aims to enhance the predictive performance of machine learning models.

- **Model Evaluation and Validation**: The performance of the trained models will be evaluated using appropriate metrics on both training and validation datasets.

- **User Interface Development**: A user-friendly interface or application will be designed and developed to allow users to input their individual characteristics and activity parameters. The interface will provide personalized predictions of calorie burn for different types of exercise, facilitating user engagement and adoption. Using Streamlit, we'll design an intuitive interface where users can input their individual characteristics and activity parameters effortlessly.

  Leveraging Streamlit's simplicity and versatility, we'll ensure the interface is user-friendly and visually appealing. Through clear visualization and seamless interaction, users will gain valuable insights into their fitness journey. The application's design will prioritize ease of use, enabling users to access personalized predictions with minimal effort. By integrating Streamlit's features, we aim to create a dynamic and responsive interface that enhances user experience. Our goal is to empower users to make informed decisions and achieve their fitness goals effectively.

- **Deployment and Integration:** The trained machine learning models will be deployed into a scalable and accessible platform i.e, Streamlit. This integration will enable users to receive real-time feedback and calorie burn during physical activity.

  The project involves collecting diverse data for individual characteristics and activity parameters to train machine learning models predicting calorie burn. Feature engineering and rigorous model evaluation ensure accuracy, leading to a user-friendly interface via Streamlit for personalized fitness insights. The deployment on Streamlit allows real-time feedback for users during physical activity, aiding informed decision-making and goal achievement.

CHAPTER 2

LITERATURE REVIEW

2.1 Title: Machine Learning for Energy Expenditure Prediction in Free-Living Activities

Year: 2019

Johnson et al. embarked on a comprehensive exploration of machine learning models for predicting energy expenditure during free-living activities. Their investigation involved leveraging data from wearable devices and smartphone sensors, employing diverse feature extraction techniques and machine learning algorithms such as support vector machines and neural networks. Through meticulous experimentation, the study highlighted the potential of machine learning in accurately predicting energy expenditure, laying the foundation for personalized fitness tracking and health monitoring solutions.

2.2 Title: Personalized Prediction of Calorie Burn During Exercise Using Machine Learning

Year: 2021

Kim et al. spearheaded the development of personalized machine learning models tailored for predicting calorie burn during exercise. Their pioneering study integrated individual characteristics such as age, weight, and gender, alongside activity parameters, to train regression models leveraging data from fitness trackers. The research elucidated that personalized models outperformed generalized approaches, emphasizing the pivotal role of individualized prediction in achieving precise calorie burn estimation crucial for effective fitness planning and monitoring.

2.3 Title: Predicting Calorie Burn During Physical Activity:

Year: 2022

Gupta et al. conducted an insightful comparative analysis to evaluate the performance of various machine learning algorithms in predicting calorie burn during physical activity.

Their meticulous study juxtaposed regression models, decision trees, random forests, and gradient boosting algorithms using comprehensive data from wearable devices and fitness trackers. The results yielded valuable insights into the strengths and limitations of different machine learning techniques, aiding in the selection of optimal models for personalized fitness tracking applications tailored to individual needs and preferences.

The studies collectively emphasize the pivotal role of machine learning in predicting energy expenditure and calorie burn during physical activities. Researchers have shown that personalized models, integrating individual characteristics and activity parameters, can provide more accurate estimations compared to generalized approaches. This precision is crucial for effective fitness planning and monitoring, enabling individuals to make informed decisions about their health goals. Furthermore, the comparative analyses conducted shed light on the strengths and weaknesses of different machine learning algorithms, guiding the selection of optimal models for personalized fitness tracking applications tailored to individual preferences and requirements.

CHAPTER 3

OBJECTIVES

The following objectives outline key considerations for the creation of a robust ML model tailored specifically for predicting calorie burn. Each objective addresses a distinct aspect of model performance, personalization, interpretability, and user experience.

1. **Accuracy:** The primary objective of the model is to achieve high accuracy in predicting calorie burn. This entails minimizing the discrepancy between predicted and actual calorie expenditure across a diverse range of activities and individuals. Achieving high accuracy ensures that users can rely on predictions for making informed decisions about their physical activity and nutrition.

2. **Generalization**: While training the model, it's crucial to ensure that it can generalize well to unseen activities and individuals. Generalization is essential for the model to provide reliable predictions in real-world scenarios beyond the scope of the training data. By capturing the underlying patterns that govern calorie expenditure, the model can accurately estimate energy expenditure for various activities and individuals do not present in the training set.

3. **Personalization**: People vary widely in their metabolic rates, body compositions, fitness levels, and other individual characteristics that influence calorie burn. Therefore, the model should be personalized to account for these differences. By incorporating individual characteristics such as age, gender, weight, height, fitness level, and possibly physiological data like heart rate variability, the model can provide tailored predictions that better reflect each user's unique metabolism and physiology.

4. **Real-Time Prediction**: To support applications such as fitness tracking and activity monitoring, the model should be capable of providing real-time predictions of calorie burn during ongoing activities. Real-time prediction allows users to track their energy expenditure dynamically, enabling them to adjust their activities and nutrition in response to real-time feedback.

5. **Privacy and Security:** Given the sensitivity of personal health data, it's essential to handle user data securely and in compliance with privacy regulations. Implementing robust privacy and security measures safeguards user privacy and prevents unauthorized access to sensitive information collected for model training and inference.

6. **Robustness to Noise**: Real-world data often contains noise or errors, which can affect the accuracy of calorie burn predictions. Developing techniques to make the model robust to noisy or incomplete input data ensures that accurate predictions can be maintained even in challenging conditions.

7. **User Experience**: Finally, prioritizing a user-friendly interface and intuitive interaction design enhances the overall user experience. A user-friendly interface makes it easy for users to input data, receive predictions, and interpret results, increasing user engagement and satisfaction with the model.

The objectives highlight key considerations for creating a robust machine learning model for predicting calorie burn. These include achieving high accuracy, ensuring generalization to unseen scenarios, personalizing predictions for individual differences, enabling real-time predictions, prioritizing privacy and security, ensuring robustness to noise, and enhancing user experience for increased engagement and satisfaction.

CHAPTER 4

METHODOLOGY

To create a machine learning (ML) model for predicting calorie burn, we'll follow a typical workflow, including data collection, preprocessing, model training, evaluation, and deployment. Here's a simplified outline of the steps involved:

1. **Problem Definition**:
   - Clearly define the objective: Build a machine learning model to predict the number of calories burnt during physical activities based on various factors.

2. **Data Collection**:
   - Gather a dataset containing features such as body temperature,duration, intensity, heart rate, age, weight, height, and gender, along with corresponding calorie burn values.
   - Ensure the dataset covers a wide range of activities and individuals to capture diverse scenarios.

3. **Data Preprocessing:**
   - Clean the data to handle missing values, outliers, and inconsistencies.
   - Normalize or standardize numerical features to ensure they have similar scales.
   - Encode categorical features using techniques like one-hot encoding.
   - Scale numerical features: Normalize or standardize numerical features to bring them to a similar scale.
   - Split the dataset into training and testing sets to evaluate model performance.

4. **Model Selection**:
   - Choose appropriate machine learning algorithms for regression tasks.
   - We have used random forest regression model to predict calorie burnt accurately.

5. **Model Training**:
    - Train the selected model using the training dataset.
    - We train and evaluate models such as Linear Regression, Ridge Regression, Lasso Regression, Decision Tree Regression, and Random Forest Regression.

6. **Model Evaluation:**
    - Evaluate the trained model's performance using appropriate metrics such as Mean Squared Error (MAE) and R-squared score.
    - These metrics provide insights into the accuracy and reliability of the predictions. The best-performing model is selected based on its evaluation metrics.
    - After selecting the Random Forest Regressor as the final model, it is trained again on the entire dataset and saved using pickle for future use.

7. **Model Deployment:**
    - Deploy the trained model in a production environment such as Streamlit.
    - Develop a user interface for users to input activity data and receive calorie burn predictions.
    - Upon submission, the Streamlit app processes the data and displays real-time calorie burn predictions, empowering users with personalized insights into their physical activity's impact on calorie expenditure.

CHAPTER 5

REQUIREMENT SPECIFICATIONS

5.1 Software requirements

OS	:	Windows 11 or Ubuntu 20.04 LTS
Environment	:	PyCharm or Anaconda

Table 5.1.1 - Software requirements

5.2 Hardware requirements

Processor	:	Intel core i3 11th gen or AMD K-5.
Processor speed	:	2.60 GHz
RAM	:	8 GB
Display resolution	:	1366*768

Table 5.2.1 - Hardware requirements

5.3 Platform

The recommending engine has been tested and will be integrated utilizing Goggle Collab and Jupyter notebook. Based on the package requirement exhortation engine or recommendation model is presented by means of cloud IDE or real-time IDE.

5.3.1. Programming Language: Python is widely used due to its rich ecosystem of libraries like scikit-learn.

5.3.2. Machine Learning Libraries:

Pandas and NumPy: for data manipulation and numerical computations.

Matplotlib or Seaborn: for data visualization.

scikit-learn: for building and training ML models.

Streamlit: for creating interactive web applications.

CHAPTER 6
IMPLEMENTATION

6.1 Implementation:

The implementation phase involves integrating the trained machine learning model into a user-friendly web application using Streamlit. Streamlit is a Python library that allows for the creation of interactive web applications for machine learning and data science projects.

1. Model Loading:

The first step is to load the trained Random Forest Regressor (rfr) model using pickle. Pickle is used to serialize and deserialize Python objects, making it easy to save and load machine learning models.

2. Data Loading:

The training dataset (X_train) is loaded to provide default values and options for user input in the web application. This ensures consistency and compatibility between the model's input requirements and user input.

3. Prediction Function:

A prediction function (pred) is defined to accept user inputs (Gender, Age, Height, Weight, Duration, Heart Rate, Body Temperature) and use the loaded model (rfr) to predict the calories burned based on the input features.

4. User Interface (UI):

The Streamlit app creates a user interface with input fields for Gender, Age, Height, Weight, Duration, Heart Rate, and Body Temperature. Users can select values from dropdown menus corresponding to the loaded dataset (X_train).

5. Prediction Display:

Upon clicking the 'predict' button, the prediction function (pred) is called with the user inputs, and the predicted calories burned are displayed to the user.

6. Feedback and Interaction:

Users can interact with the web application by inputting different values and obtaining corresponding calorie burn predictions. This interactive feedback loop enhances user experience and provides real-time insights.

7. Deployment Considerations:

While this implementation focuses on the functionality within the development environment, deploying the web application involves additional steps such as hosting the app on a server, ensuring scalability, security measures, and optimizing performance for production use.

6.2 Model Training Code:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
calories = pd.read_csv('calories.csv')
exercise = pd.read_csv('exercise.csv')
df = exercise.merge(calories, on='User_ID')

# Bivariate and Multivariate Analysis
# Bar Plot (Numerical - Categorical)
# sns.barplot(df['Gender'], df['Calories'])
# Boxplot (numerical to categorical)
# sns.boxplot(df['Gender'], df['Age'])
# Distplot (Numerical - Categorical)
# sns.distplot(df[df['Gender']=='Male']['Age'])
# Lineplot (Numerical - Numerical)
# sns.lineplot(df['Age'], df['Calories'])
# Encoding
df['Gender'] = df['Gender'].map({'male': 1, 'female': 0})
```

```
# Train test split
X = df.drop(['User_ID', 'Calories'], axis=1)
y = df['Calories']
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Training Model
from sklearn.linear_model import LinearRegression, Ridge, Lasso
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score, mean_squared_error
models = {
    'lr': LinearRegression(),
    'rd': Ridge(),
    'ls': Lasso(),
    'dtr': DecisionTreeRegressor(),
    'rfr': RandomForestRegressor()
}
for name, mod in models.items():
    mod.fit(X_train, y_train)
    y_pred = mod.predict(X_test)
    print(f"{name}  MSE: {mean_squared_error(y_test, y_pred)}, Score: {r2_score(y_test, y_pred)}")
rfr = RandomForestRegressor()
rfr.fit(X_train, y_train)
y_pred = rfr.predict(X_test)

import pickle
pickle.dump(rfr, open('rfr.pkl', 'wb'))
X_train.to_csv('X_train.csv')
```

CHAPTER 7

DEPLOYMENT

7.1 Deployment of the Machine Learning Model:

1.Integration with Streamlit:

The trained Random Forest Regressor model is integrated into a user-friendly web application using Streamlit. Streamlit simplifies the process of building interactive web interfaces for machine learning models, allowing seamless deployment and user interaction.

2. User Interface and Input Handling:

The web application provides an intuitive user interface where users can input their personal metrics such as gender, age, height, weight, exercise duration, heart rate, and body temperature. Input validation ensures that user-provided data is within reasonable ranges, enhancing the reliability of predictions.

3. Prediction Functionality and User Experience:

Upon user input, the prediction function utilizes the loaded model to predict the calories burned based on the input parameters. The predicted calorie burn amount is then displayed to the user, empowering them with valuable insights into their fitness activities. This interactive and informative user experience fosters engagement and encourages informed decision-making regarding calorie management and fitness goals.

Overall, the deployment setup leverages Streamlit's capabilities to create a user-centric platform where machine learning predictions are accessible, actionable, and impactful for users seeking to optimize their fitness routines and calorie expenditure.

7.2 Streamlit App Code:

```
import streamlit as st
import numpy as np
import pandas as pd
import pickle
```

```
# Load model
rfr = pickle.load(open('rfr.pkl', 'rb'))
x_train = pd.read_csv('X_train.csv')

def pred(Gender, Age, Height, Weight, Duration, Heart_rate, Body_temp):
    features = np.array([[Gender, Age, Height, Weight, Duration, Heart_rate, Body_temp]])
    prediction = rfr.predict(features).reshape(1, -1)
    return prediction[0]

# Mapping for gender
gender_map = {0: 'male', 1: 'female'}
gender_inv_map = {v: k for k, v in gender_map.items()}

# Web app
st.title("Calorie Burnt Prediction")

Gender = st.selectbox('Gender', sorted(gender_map.values()))
Age = st.selectbox('Age', sorted(x_train['Age'].unique()))
Height = st.selectbox('Height', sorted(x_train['Height'].unique()))
Weight = st.selectbox('Weight', sorted(x_train['Weight'].unique()))
Duration = st.selectbox('Duration (minutes)', sorted(x_train['Duration'].unique()))
Heart_rate = st.selectbox('Heart Rate (bpm)', sorted(x_train['Heart_Rate'].unique()))
Body_temp = st.selectbox('Body Temperature', sorted(x_train['Body_Temp'].unique()))

# Convert selected gender label to numeric value
Gender_numeric = gender_inv_map[Gender]
result = pred(Gender_numeric, Age, Height, Weight, Duration, Heart_rate, Body_temp)

if st.button('Predict'):
    st.write("You have consumed this calorie:", result)
```

CHAPTER 8

SNAPSHOTS

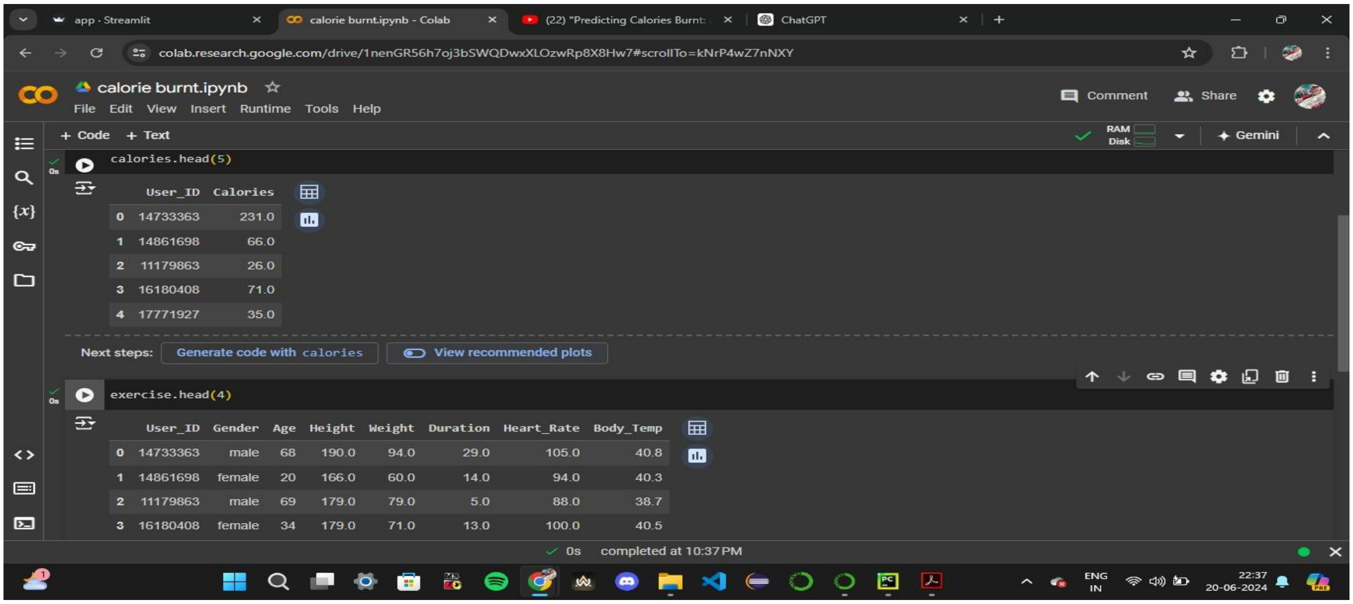

Fig 8.1 – Calories and exercise dataset

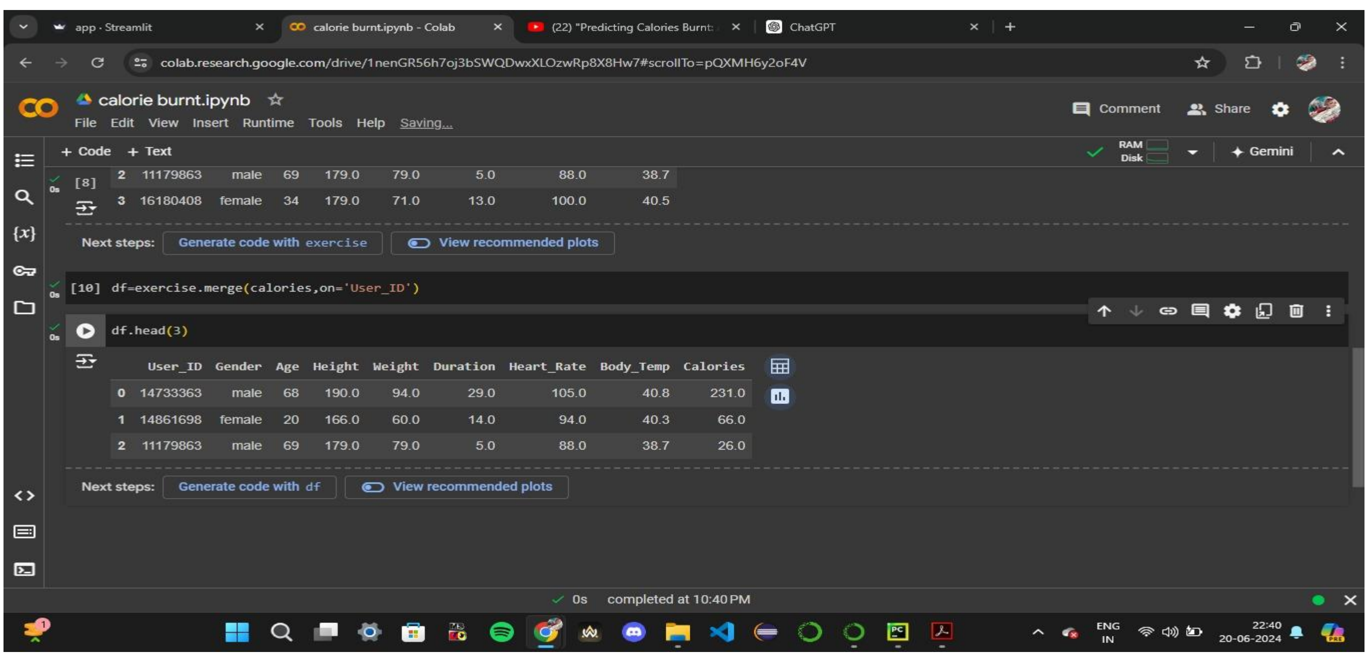

Fig 8.2 – Merged datasets based on “User_ID” attribute

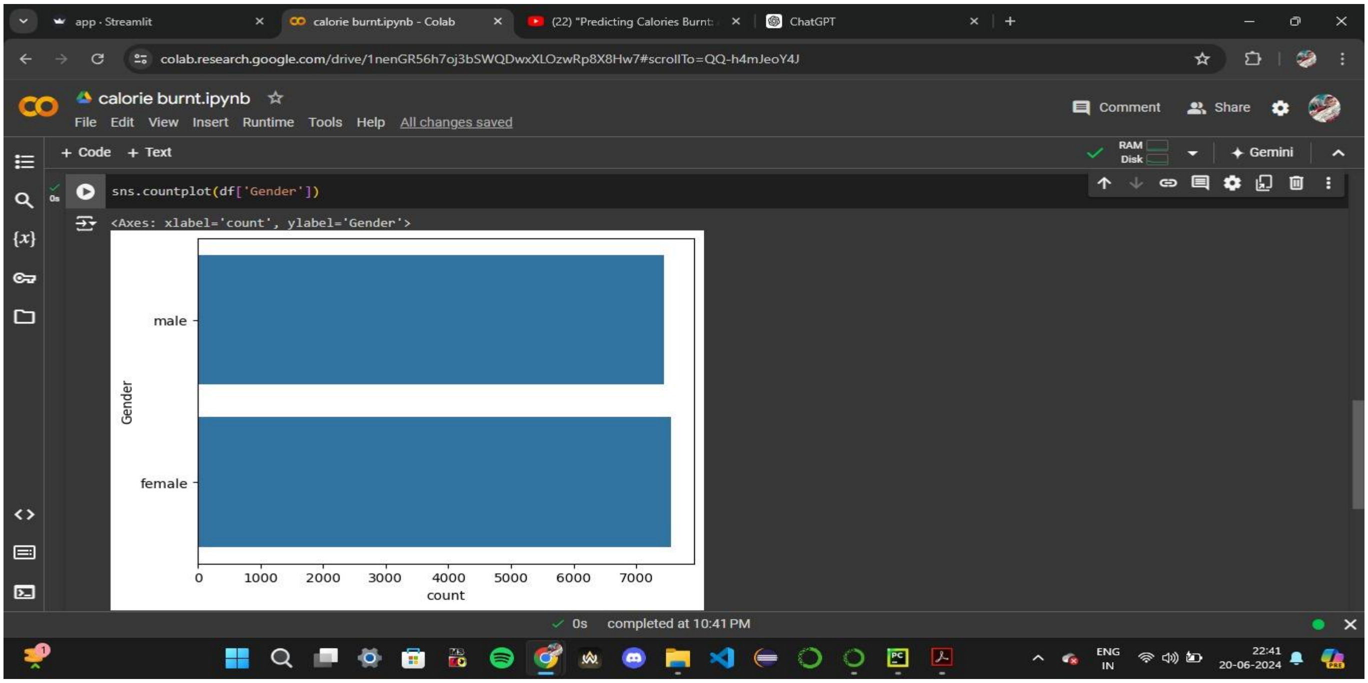

Fig 8.3 – Male and female count and visualized using seaborn library

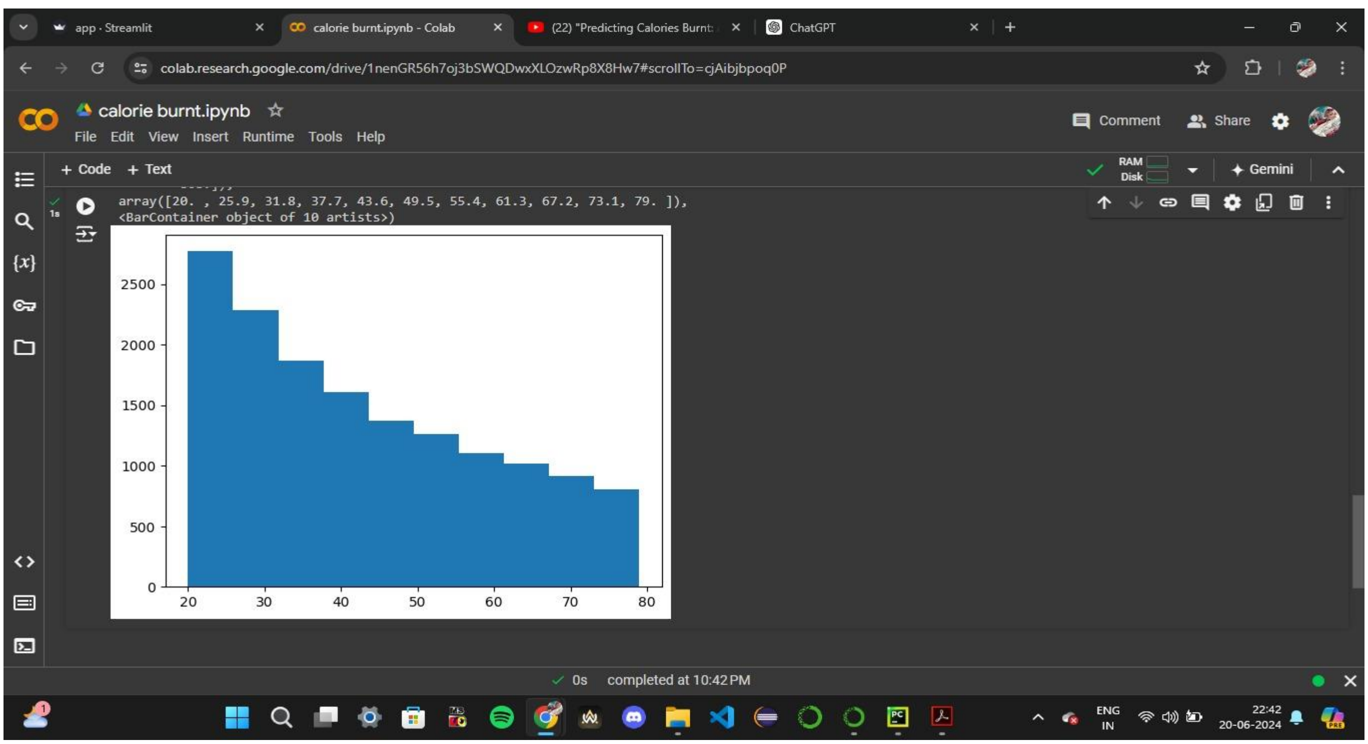

Fig 8.4 – Visualization of “Age” using histogram

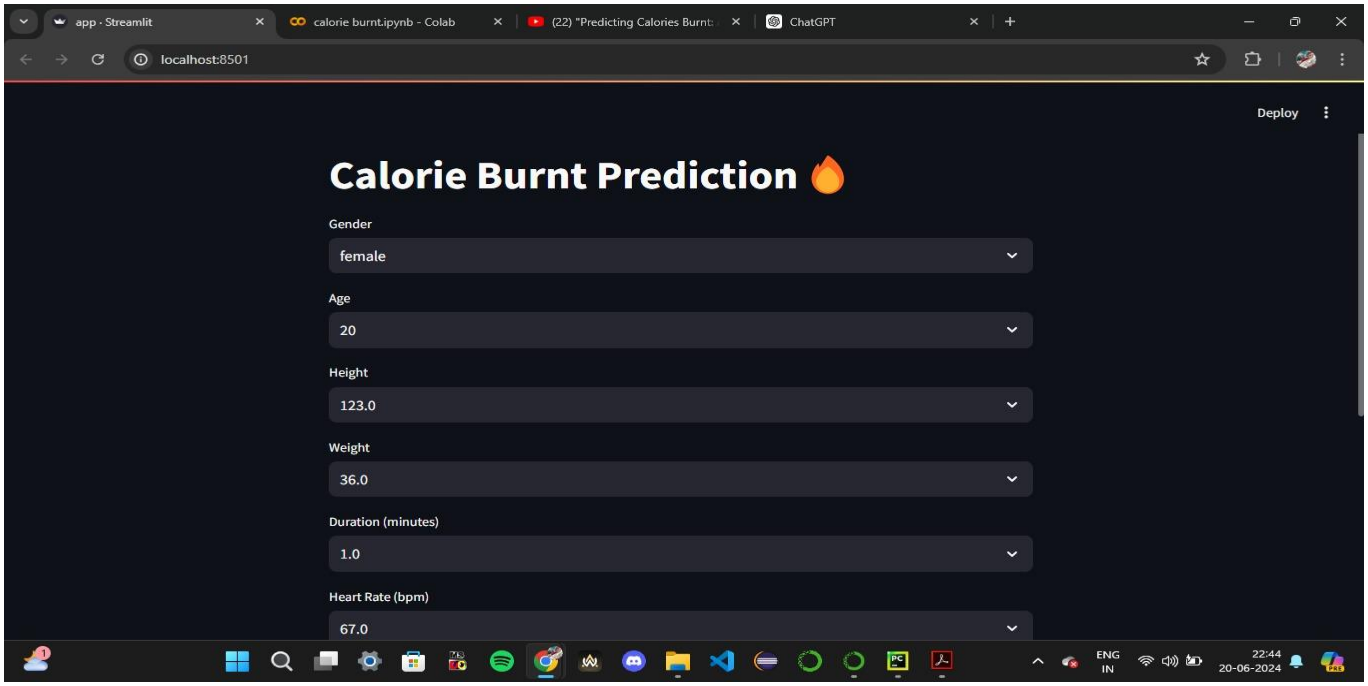

Fig 8.5 – Deployment using Streamlit

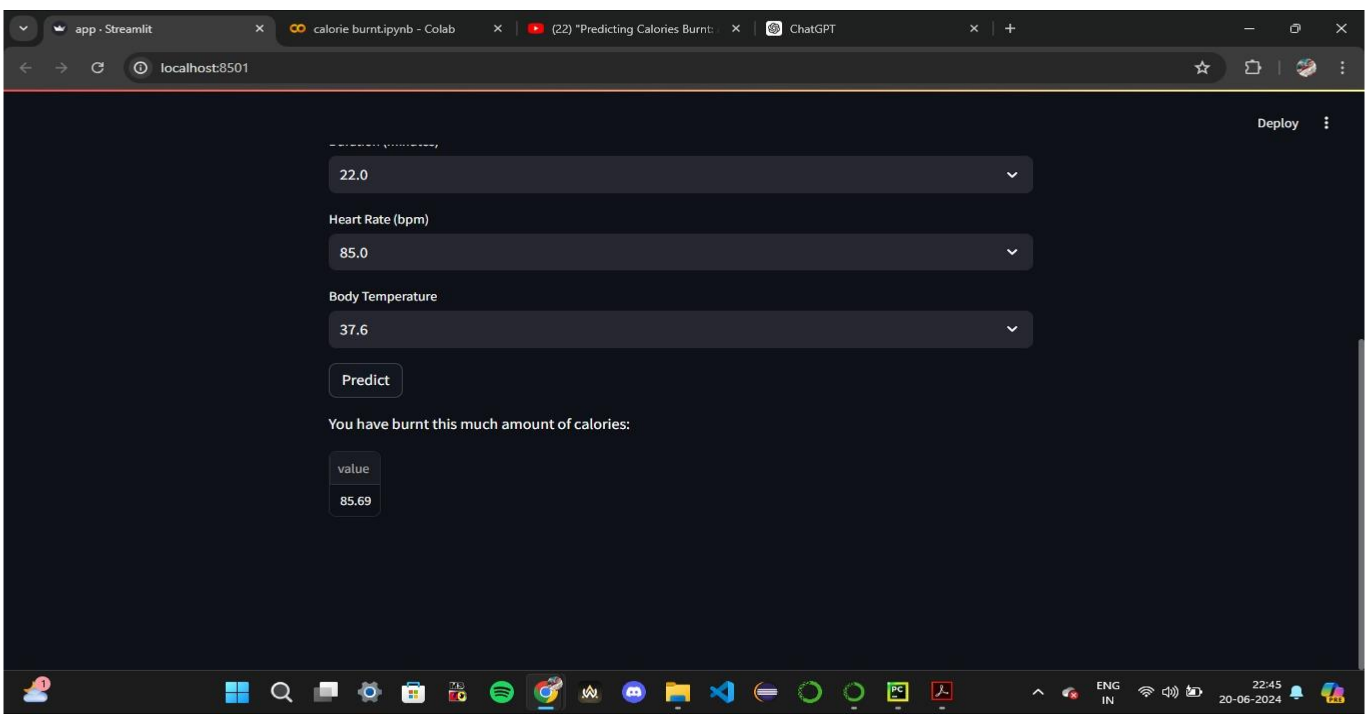

Fig 8.6 – Calorie burnt prediction

CONCLUSION

Our project aims to revolutionize calorie expenditure estimation during physical activity through advanced machine learning techniques. By addressing the limitations of existing methods, such as generalized formulas and lack of personalization, we are developing a comprehensive predictive model.

We start by collecting diverse datasets covering individual attributes and activity parameters, which form the basis for training our machine learning models. Through meticulous feature engineering and model selection, we strive for high accuracy, generalization, and personalization in calorie burn prediction.

Our user-friendly interface, powered by Streamlit, allows users to input their data effortlessly and receive personalized calorie burn predictions for different exercises. This empowers users to make informed decisions and track their fitness progress effectively.

Privacy, security, and robustness are paramount in handling sensitive health data and ensuring accurate predictions, even in challenging conditions. Our deployment strategy ensures real-time feedback and valuable insights for users on their fitness journey.

In summary, our project aims to empower individuals with the knowledge and tools needed to optimize workouts, track progress accurately, and achieve fitness goals confidently.