Smart Internz

Anticipating Caloric Expenditure With Machine Learning

Short Term Internship

PROJECT REPORT

Team Members:

1. John wesely
2. Narasinga Reddy
3. Tulasi Ram
4. Shabeer

Introduction

Predicting Calorie Consumption is a Crucial aspect of Nutrition and health management. By using data such as age, gender weight, activity level and dietary choices, predictive models can estimate daily calorie needs.

These predictions aid individuals in making informed decisions about their diet and can be instrumental in acheiving health and fitness goals. In this context, we provide data set and train our model to predict How many calories Burned during Exercises

1. Data Collection

Gather a diverse dataset that includes information like age, gender, weight, Height, duration, Heart rate and Body temperature during exercise

2. Data Preprocessing

- Clean the data by Handling missing values and Outliers
- Normalize or standardize Numerical Features

3. Model Selection

- choose appropriate Machine learning algorithms for regress...

such as linear regression, decision trees, random forests, or Neural Networks

4. Training and Validation

- Split the dataset into training and testing sets
- Train the selected models on the training data and evaluate their performance using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Rsquared.

5. Model Evaluation.

- Assess the models generalization on the testing data

6. Deployment:

- Once a satisfactory model is Achieved, deploy it as an application or service where users can input their details and get Calorie Expenditure prediction.

DESCRIPTION

This project focuses on developing a predictive model to estimate the No. of calories a person Burns during various physical activities. The primary goal of this project is to assist individuals in managing their fitness and health. The project could lead to the development of a user-friendly

application or Service that allows users to input their information and get real time calorie expenditure predictions.

This can be a valuable Resource for people aiming to make informed Decisions about their diet and Exercise routines, ultimately promoting a Healthier Lifestyle

PURPOSE

1. Health and Fitness Management:

The project aims to assist individuals in managing their health and fitness more effectively by providing accurate estimates of the calories burned during various physical Activities. This information can help people make informed decisions about their exercise routines and dietary choices

2. Data-Driven Decision Making:

The project leverages machine learning to provide data driven insights. It emphasizes the importance of using technology to make informed choices regarding calorie intake and expenditure, aligning with modern trends in health and fitness

LITERATURE SURVEY

In a calorie expenditure prediction project, Several existing approaches and methods can be used to address the problem effectively. Here are some common approcades

1. Linear Regression:

- Linear regression models can be employed to establish a linear relationships Between input features (such as age, weight, activity duration) and the output variable (calorie expenditure).

2. Decision Trees and Random Forests:

Decision trees and random forests are useful for capturing non linear relationships and interactions Between features they can handle Compler feature interactions and provide interpretability.

3. XG Boost Regressor

XG Boost is a powerful gradient Boosting framework that is widely used in machine learning for regression tasks.

My Choice

Using the XG Boost regressor is an excellent choice for our calorie expenditure prediction project. Here are some reasons why the XG Boost regressor is strong candidate for our project

1. High Predictive Accuracy
2. Handling Non-Linearity between input features and Output Variable
3. Feature Importance
4. Scalability
5. Tuning options of Hyper parameters

The XG Boost regressor is a robust choice, offering Both accuracy and interpretability, making it a popular method for regression tasks in Machine Learning.

Hardware and Software requirements of the Project

Hardware Required : System or Laptop

Software Required : Anaconda Navigator, Flask, Python, XgBoost library, streamlit

System Required : Windows (7, 8, 9, 10, 11), 4 GB RAM, 256 GB Hard Disk.

4. EXPERIMENTAL INVESTIGATIONS

Here are key aspects of experimental investigations for such a project predicting calorie expenditure

1. Data splitting:

We divided the given dataset into two parts, typically training and testing sets

2. Model Training:

We used XG Boost regressor to train the model and the dataset.

3. Model Evaluation:

We Evaluated the model using appropriate regression metrics such as Mean Absolute Error (MAE),

and R-Squared (R_2) error

Comparative Analysis:

we have explored multiple Machine Learning algorithms (eg. XGBoost, linear regression etc), compare their performance to identify the most effective approach

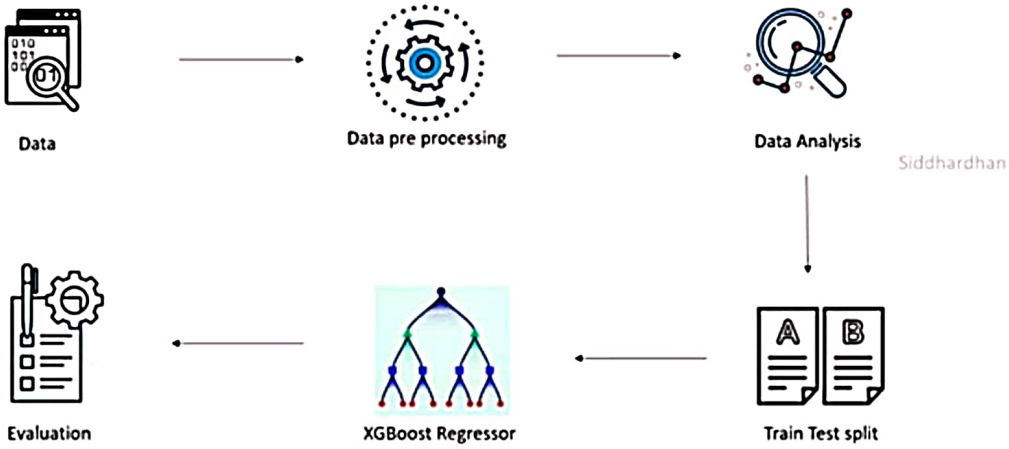
Model Visualization:

we Visualized the models predictions and compared them to actual values. Visulization Helped in understanding the models Behavior and any Potential Shortcomings

Feature Importance Analysis

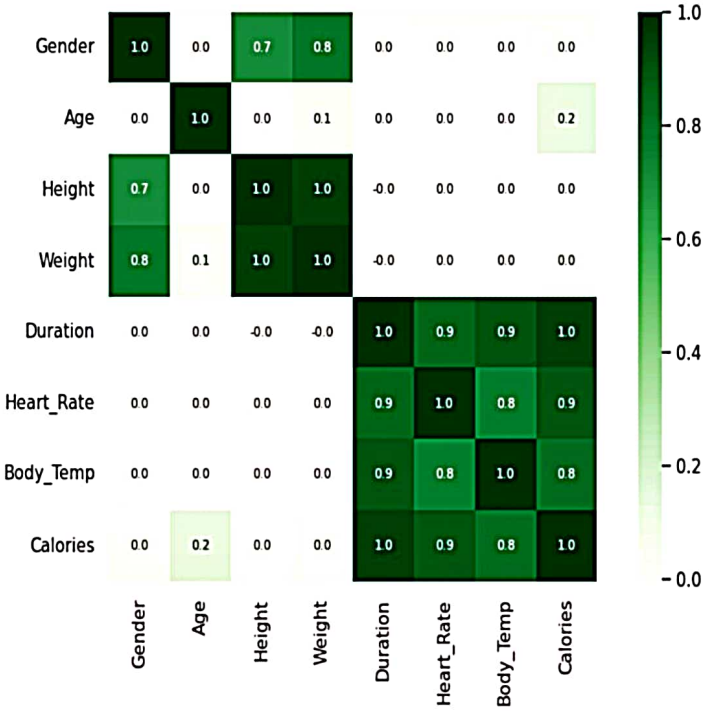
We have Understood which features (age, weight, activity duration, etc) have most significant impact on calorie expenditure. Feature importance analysis helps in Model interpretation

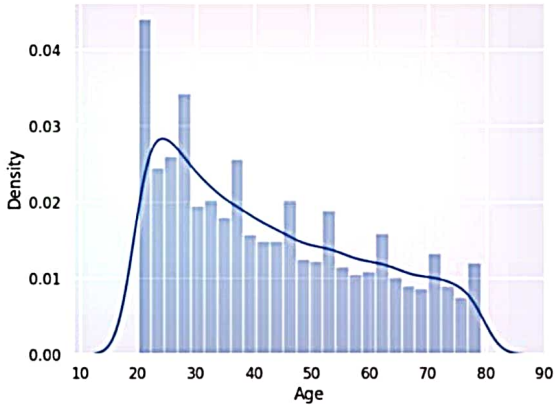
Work Flow

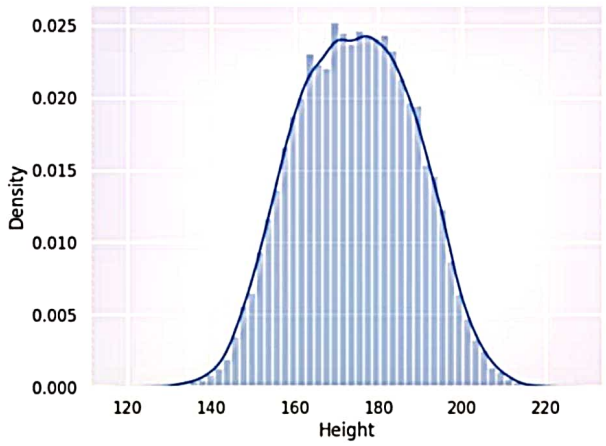


Graphs

Graphs







ADVANTAGES

1. Health and Fitness Management:

The project can empower individuals to make informed decisions about their health and fitness, promoting a healthier lifestyle

2. Customized Recommendations:

It provides personalized recommendations for calorie expenditure Based on Individual characteristics and activity levels

3. Predictive Accuracy:

Machine Learning models can provide reasonably accurate predictions of calorie expenditure, especially when fine-tuned.

4. Continuous Learning:

The project can continuously improve its accuracy by incorporating new data and research findings.

5. Accessibility:

user-friendly applications or services make this information accessible to a wide Audience

DISADVANTAGES

1. **Data Privacy:**

Handling personal health and fitness data requires careful data privacy measures to protect user information

2. **Data Quality:**

The Accuracy of Predictions is Heavily dependent on the Quality and representativeness of the dataset

Inaccurate or Biased data can lead to poor Predictions

3. **Model Complexity:**

more accurate models, such as deep neural Network, can be computationally expensive and require Substantial data

4. **Resource Intensive:**

Training and maintaining machine learning models can require significant Computational resources, particularly for deep Learning models

5. **Dependency on Data Updates:**

The Accuracy of the System relies on regular update with new data and research findings

APPLICATIONS

There are various practical applications in health, fitness and wellness domains. Here are some key applications

1. personal Fitness and Health Management
2. Diet and Nutrition planning
3. Weight Loss programs
4. Wearable Fitness Devices
5. Healthcare and Rehabilitation
6. Research Studies
7. Food and Beverage Industry
8. Fitness Apps and Services
9. Corporate Wellness Programs
10. Public Health Initiatives

CONCLUSION:

A Calorie Expenditure prediction Project is a valuable initiative that combines the power of machine learning and data analysis to offer personalized insights into an individuals calorie Burn during physical activities Calorie Burn during physical Activities

this project has the potential to revolutionize How people Manage their Health and Fitness

FUTURE SCOPE

The field of Calorie Expenditure prediction is dynamic. there are Numerous Future enhancements and developments that can be made to further improve the accuracy, usability and impact of Such projects.

1. Real-time Monitoring:

Develop real-time Monitoring Solutions that continuously track Calorie expenditure during Various activities

2. Advanced Data Sources:

Incorporate more diverse and advanced data Sources, such as Biometric Sensors. GPS, and metabolic rate measurements to enhance prediction Accuracy

3. Personalized Meal Planning:

Integrate calorie expenditure predictions with dietary planning to provide holistic guidance for users, considering both energy expenditure and intake.

4. Behavioral Insights:

Analyze user behaviour data to provide insights into patterns and habits, helping individuals make long-term lifestyle changes

5- Machine Learning Advancements:

. Explore the use of advanced machine learning techniques such as reinforcement learning, to optimize individualized exercise routines

The future of calorie expenditure prediction projects is exciting, with opportunities for innovation and positive impact on individual's Health and Well-Being. By staying attuned to emerging technologies and research findings, these projects can continue to evolve and offer more comprehensive accurate and user friendly solutions

BIBLIOGRAPHY

1. Reference Books:-

"Machine learning Yearning" by Andrew Ng

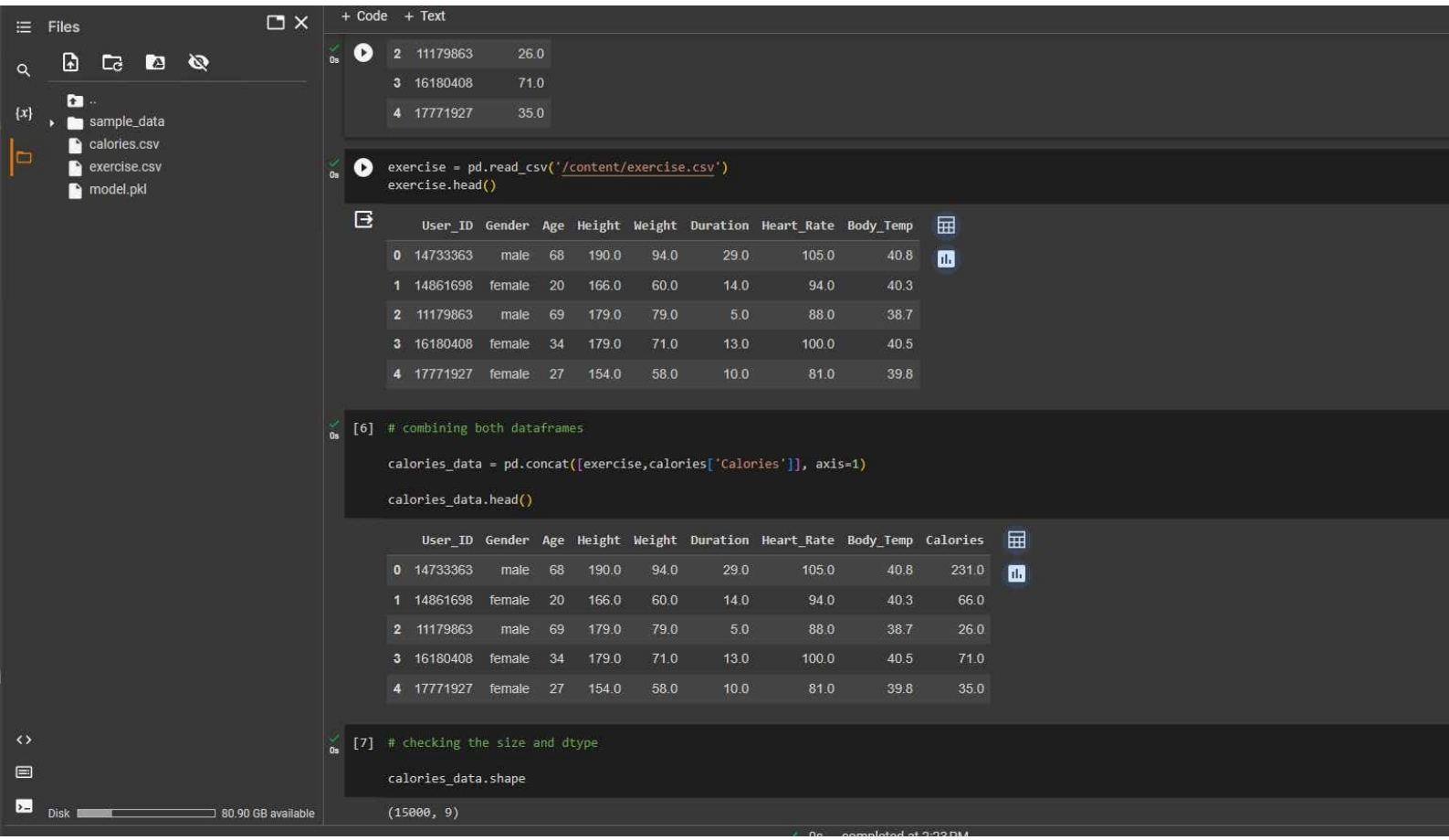
2. "Nutrition for Sport and Exercise" by Marie Dunford

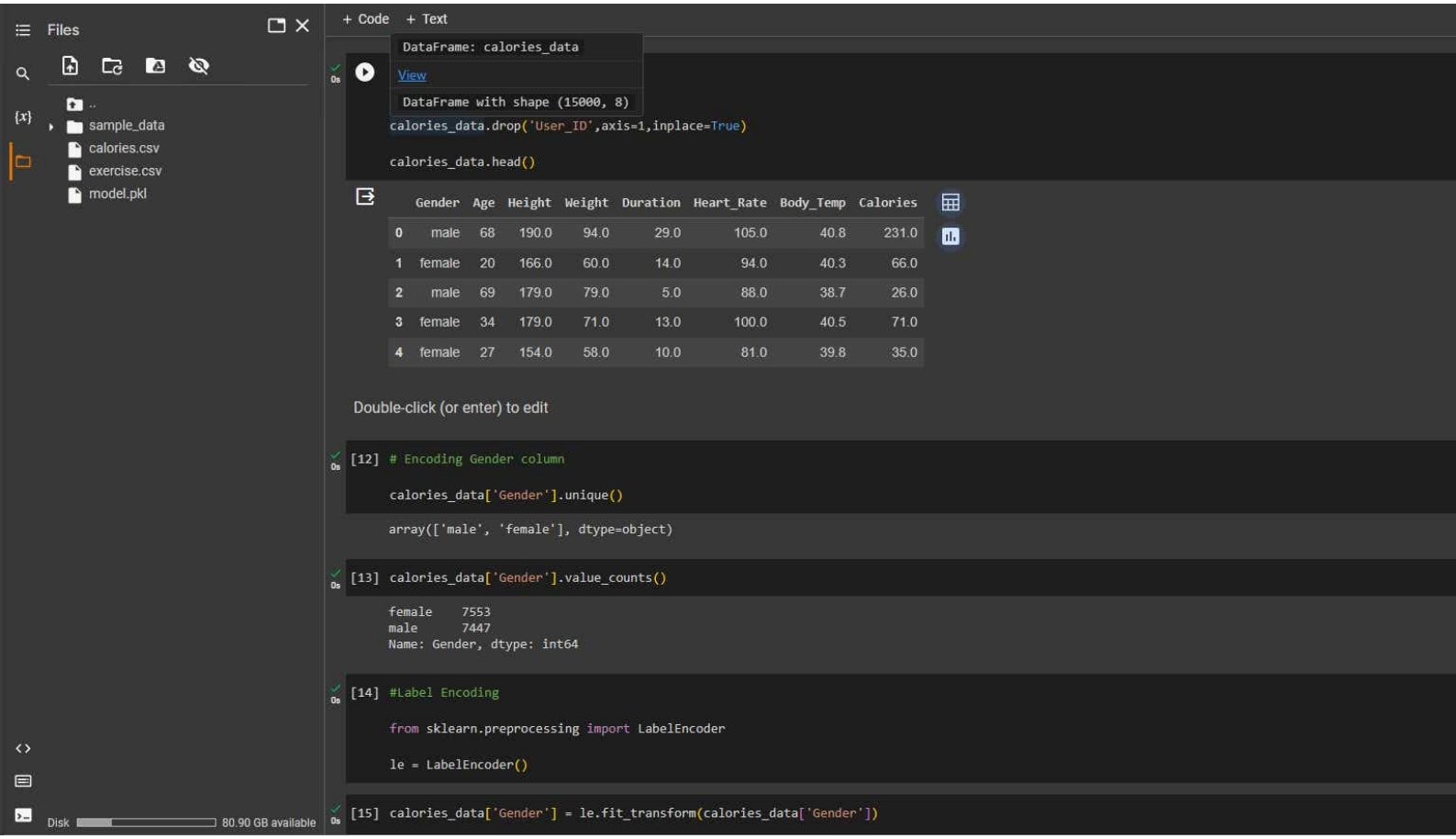
3. "Data Science for Health Care:
Methodologies and Applications" by
Subhash C. Basak and Ronau P. Singh

Reference Websites:

GeekforGeeks, Github and kaggle

APPENDIX


Files
sample_data
calories.csv
exercise.csv
model.pkl
Disk
80.90 GB available
+ Code
+ Text
2 11179863 26.0
3 16180408 71.0
4 17771927 35.0
exercise = pd.read_csv('/content/exercise.csv')
exercise.head()
User_ID Gender Age Height Weight Duration Heart_Rate Body_Temp
0 14733363 male 68 190.0 94.0 29.0 105.0 40.8
1 14861698 female 20 166.0 60.0 14.0 94.0 40.3
2 11179863 male 69 179.0 79.0 5.0 88.0 38.7
3 16180408 female 34 179.0 71.0 13.0 100.0 40.5
4 17771927 female 27 154.0 58.0 10.0 81.0 39.8
[6] # combining both dataframes
calories_data = pd.concat([exercise,calories['Calories']], axis=1)
calories_data.head()
User_ID Gender Age Height Weight Duration Heart_Rate Body_Temp Calories
0 14733363 male 68 190.0 94.0 29.0 105.0 40.8 231.0
1 14861698 female 20 166.0 60.0 14.0 94.0 40.3 66.0
2 11179863 male 69 179.0 79.0 5.0 88.0 38.7 26.0
3 16180408 female 34 179.0 71.0 13.0 100.0 40.5 71.0
4 17771927 female 27 154.0 58.0 10.0 81.0 39.8 35.0
[7] # checking the size and dtype
calories_data.shape
(15000, 9)


Files
..
sample_data
calories.csv
exercise.csv
model.pkl
Disk
80.90 GB available
+ Code
+ Text
DataFrame: calories_data
View
DataFrame with shape (15000, 8)
calories_data.drop('User_ID',axis=1,inplace=True)
calories_data.head()
Gender Age Height Weight Duration Heart_Rate Body_Temp Calories
0 male 68 190.0 94.0 29.0 105.0 40.8 231.0
1 female 20 166.0 60.0 14.0 94.0 40.3 66.0
2 male 69 179.0 79.0 5.0 88.0 38.7 26.0
3 female 34 179.0 71.0 13.0 100.0 40.5 71.0
4 female 27 154.0 58.0 10.0 81.0 39.8 35.0
Double-click (or enter) to edit
[12] # Encoding Gender column
calories_data['Gender'].unique()
array(['male', 'female'], dtype=object)
[13] calories_data['Gender'].value_counts()
female 7553
male 7447
Name: Gender, dtype: int64
[14] #Label Encoding
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
[15] calories_data['Gender'] = le.fit_transform(calories_data['Gender'])

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


model = pickle.load(open('./model/model.pkl','rb'))

gender = [0,1]

st.title('Calorie Prediction App')

Gender = st.selectbox('Gender ( 0: female, 1: male)',gender)

col1,col2,col6 = st.columns(3)
with col1:
    Age = st.number_input('Age')
with col2:
    Height = st.number_input('Height')
with col6:
    Weight = st.number_input('Weight')

col3,col4,col5 = st.columns(3)
with col3:
    Duration = st.number_input('Exercise duration (in mins)')
with col4:
    Heart_Rate = st.number_input('Heart Rate')
with col5:
    Body_Temp = st.number_input('Body Temp (in Celsius)')

if st.button('Predict'):
    input_data = np.array([Gender,Age,Height,Weight,Duration,Heart_Rate,Body_Temp])
    input_df = input_data.reshape(1,-1)
    result = model.predict(input_df)
    st.header(f'result is :  {int(result[0])}')
```