Regression Models to Predict Human Body Fat

Team: Plantae
Student name: Yufan Zhang & Ai-Shuan Lee & Amber Ye
ANS128 3/13/2025

Table of Contents

Abstract

Summary of Project

The project focuses on predicting human body fat percentage using regression models. This project developed a more accurate Body Fat Calculator using body measurement.

Why is it important?

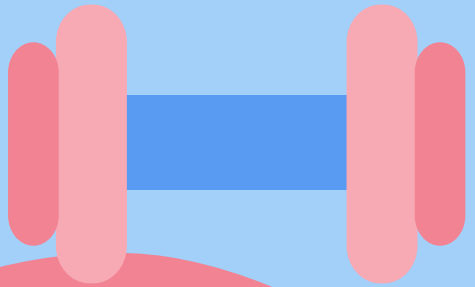
Too much or less body fat can increase the risk of some serious health issues.

What was done?

Various regression models has been tested, and a Body Fat Calculator was built using the most correlated body traits.

Key findings & takeaways

The new model is more accurate than other ways.

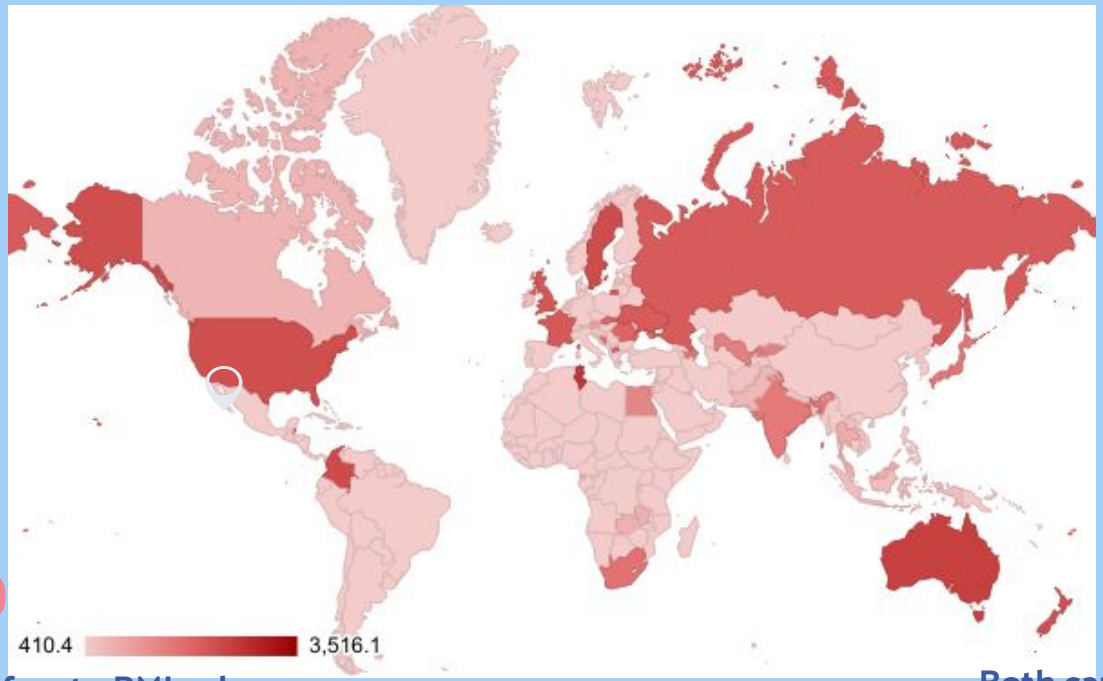


01
Introduction & Problem Statement

Introduction

60%

In United States, it's estimated that over 60% of adults are overweight or obese.



40%

Globally, 40% adults were overweight or obese in 2016.

BMI 25–30

body-mass index – it refers to BMI values between 25 and 30.

High or low body fat?

Both can lead to health issues.
High : heart disease, type 2 diabetic
Low: weakened immune system, hormonal imbalances.

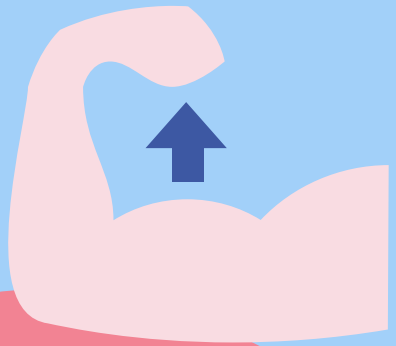
Problem Statement

Objective

Traditionally, people like to using body mass index (BMI) to define obesity or skinny. However, studies have shown that the traditional method of predicting body fat percentage using body mass index (BMI) can be misleading and inaccurate(Woolcott,2023).

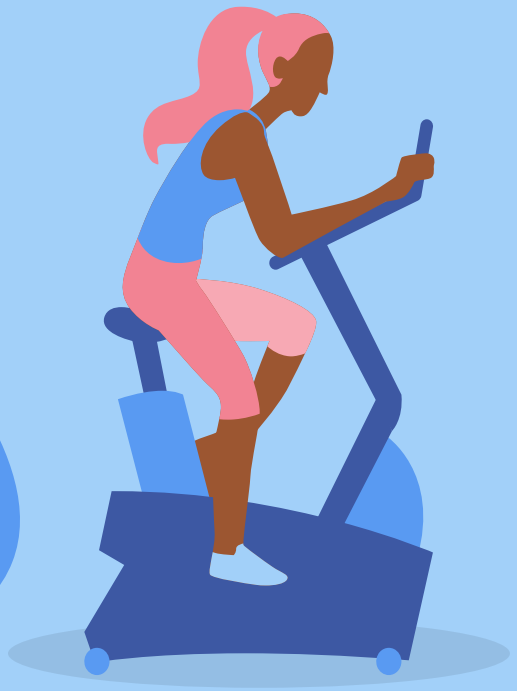
This project aims to develop a more accurate and accessible model for predicting body fat percentage using body measurement.

Relevance to Biology.



02

Literature Review



Literature Review

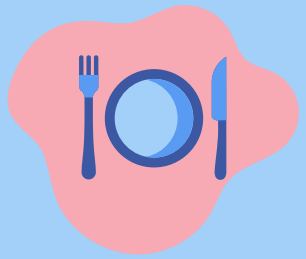
BMI is widely used but often inaccurate for body fat estimation. Woolcott and Bergman (2018) proposed Relative Fat Mass (RFM), which better predicts body fat using height and waist measurements.

However, RFM needs validation across different populations, and a universal standard for obesity based on body fat is still lacking.

This project set up a more specific body fat predict model different with BMI and RFM.

03

Dataset Description, Analysis & Visualization



Dataset Description

Basic Information & Details

- From Body Fat Prediction Extended on Kaggle
- 16 anthropometric traits from 252 male and 184 female
- Dr. A Garth Fisher

Key Features

1. Body Fat(%)
2. Original(Y)
3. Sex (M/F)
4. Age(years)
5. Weight(kg)
6. Height (mt)
7. Neck (cm)
8. Chest (cm)
9. Abdomen (cm)
10. Hip (cm)
11. Thigh (cm)
12. Knee (cm)
13. Ankle (cm)
14. Biceps (cm)
15. Forearm (cm)
16. Waist (cm)

Data Processing

- **No Missing Values**
- **Handling Outliers**
  - Outliers: can sometimes skew the results of analysis.
- **Replace 'Gender' and remove 'Original'**

Data Analysis

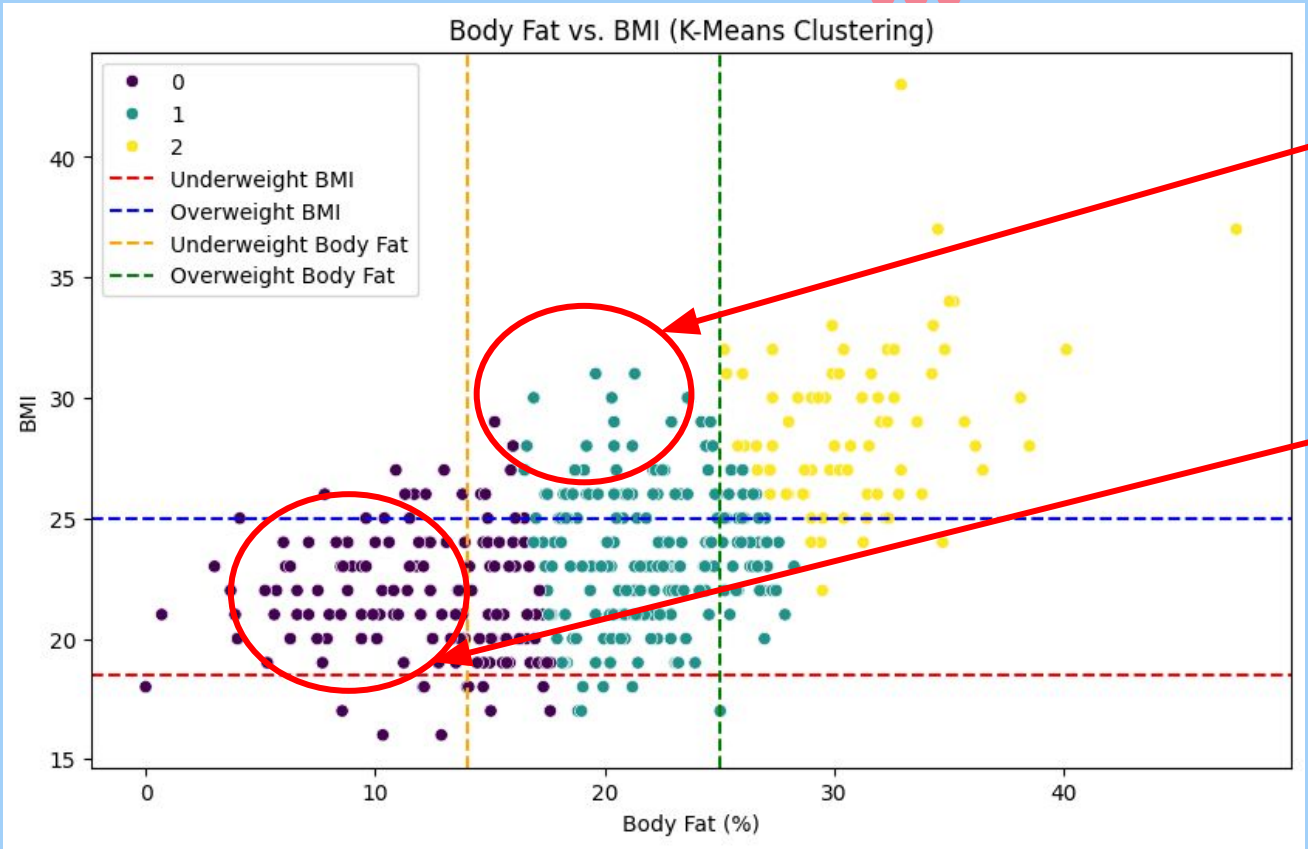
1. Including BMI

- BMI estimates Body Fat
- Weight (kg) /Height $(m)^2$
- Include in Original dataset

2. Compare BMI and Body Fat %

- Underweight: BMI Value <18.5, Body Fat <14%
- Normal: 18.5 <BMI Value<25, 14% <Body Fat<25%
- Overweight: BMI Value >25, Body Fat >25%
- 70 individuals wrongly categorized

Data Analysis (K-cluster)



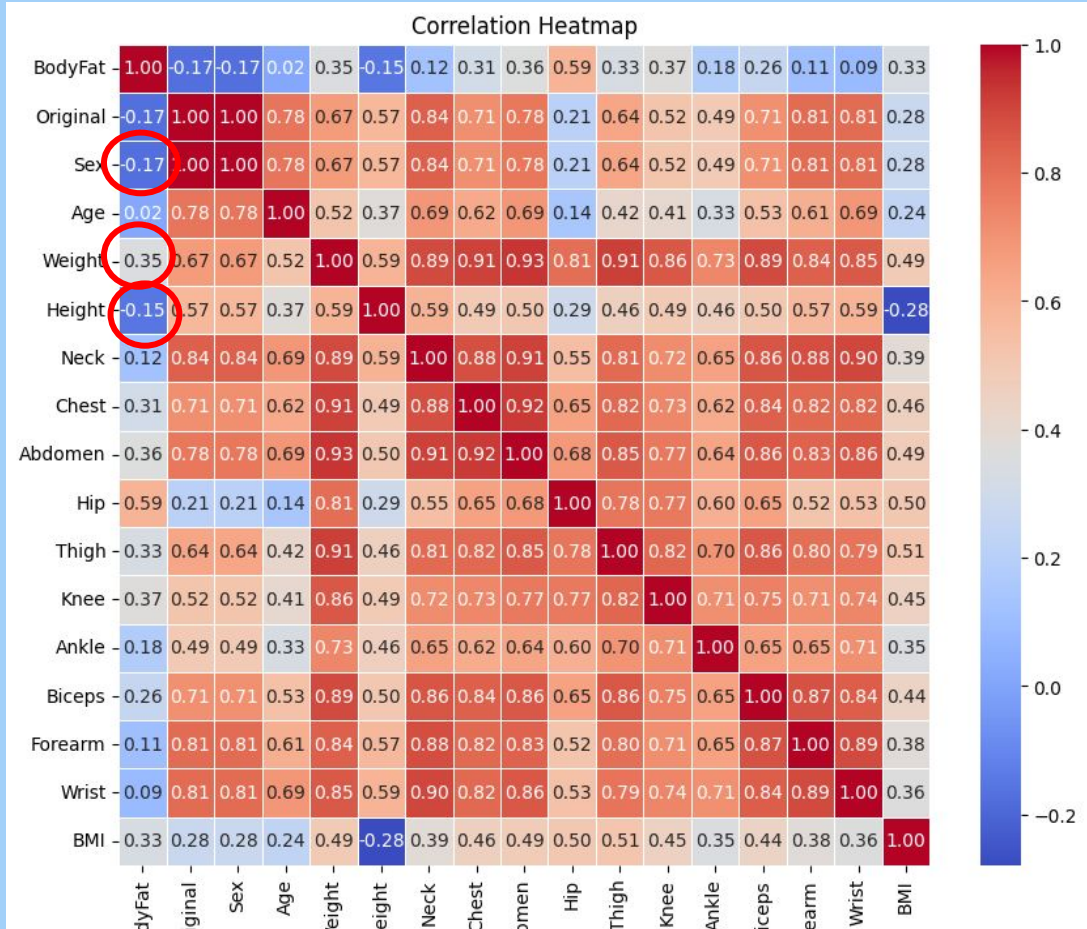
3. Disparity

Normal bodyfat percentage but BMI is overweight

Low Body Fat Percentage but normal BMI

Initial surprising findings

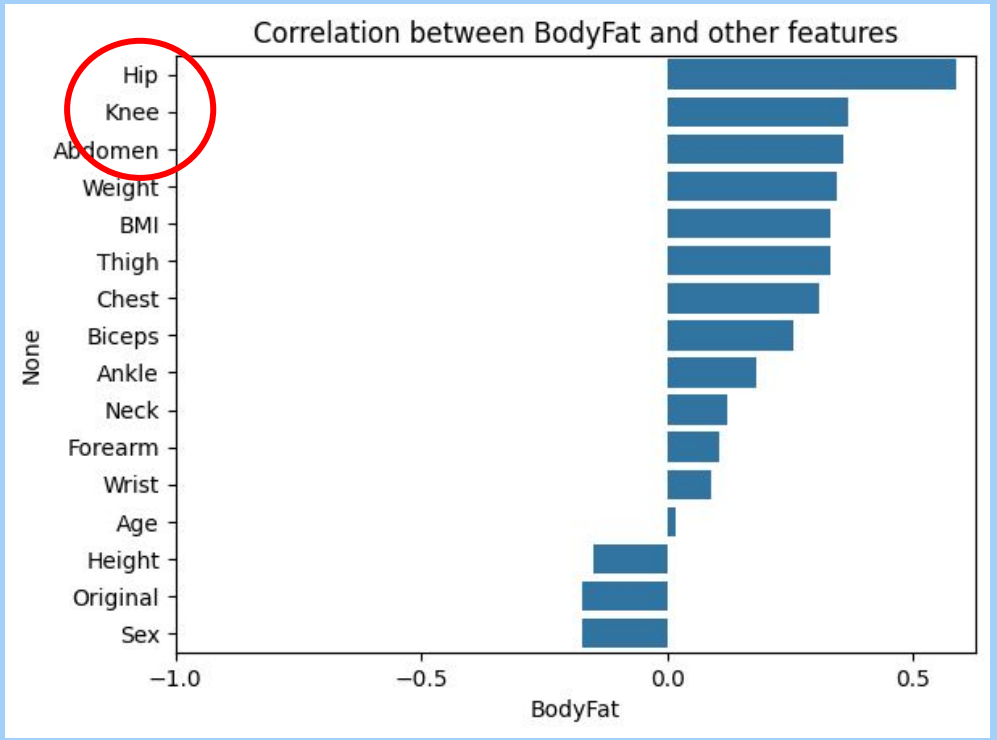
- Expectation: Gender, Height and Weight to be most correlated
- Correlation Heatmap



Data Analysis and Visualization

Correlation Among Each Traits

- Correlation Bar Chart between Body Fat% and other traits
- “Hip”, “Knee”, “Abdomen” circumference



Linear Regression Model

Variables

- Hip circumference
- Knee circumference
- Abdomen circumference

Linear Regression

- Coefficients: [0.930495, -0.89346946, 0.03874588]
- Intercept: -41.27719724375874
- Mean Squared Error (MSE): 32.496548955682826
- Root Mean Squared Error (RMSE): 5.700574440850924
- R-squared (R^2) 0.2977274240809725

Final Equation

- -41.28+0.93X-0.89Y+0.039Z
- X=hip circ.
- Y=knee circ.
- Z=abdomen circ.

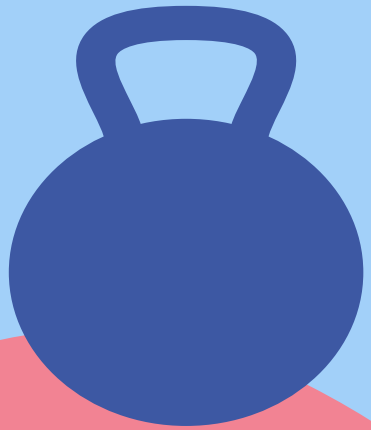
Linear Regression Model

Adding More Variables

- Weight, Thigh, Chest
- Total 6 variables

R^2 Comparison

- 3 variables ~0.2861
- 6 variables ~ 0.4341



04

Model Development & Evaluation

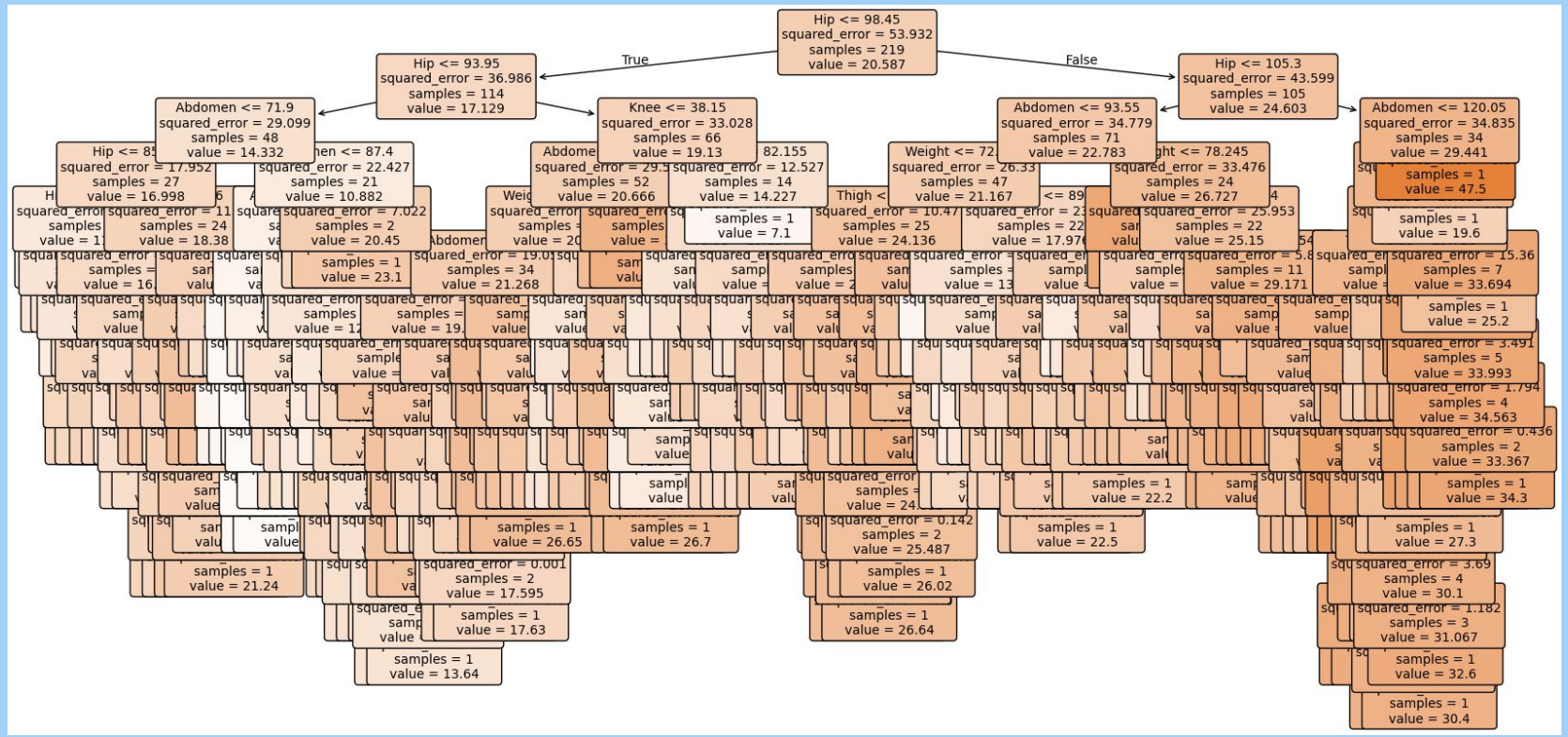
Comparing Different Models

- Linear Regression, Random Forest Regression, and Support Vector Regression.

	Linear Regression	Random Forest	Support Vector
MSE	32.4965	26.1378	34.4417
RMSE	5.7006	5.1125	5.8687
R^2	0.2977	0.4351	0.2557

However...

Random Forest Regressor

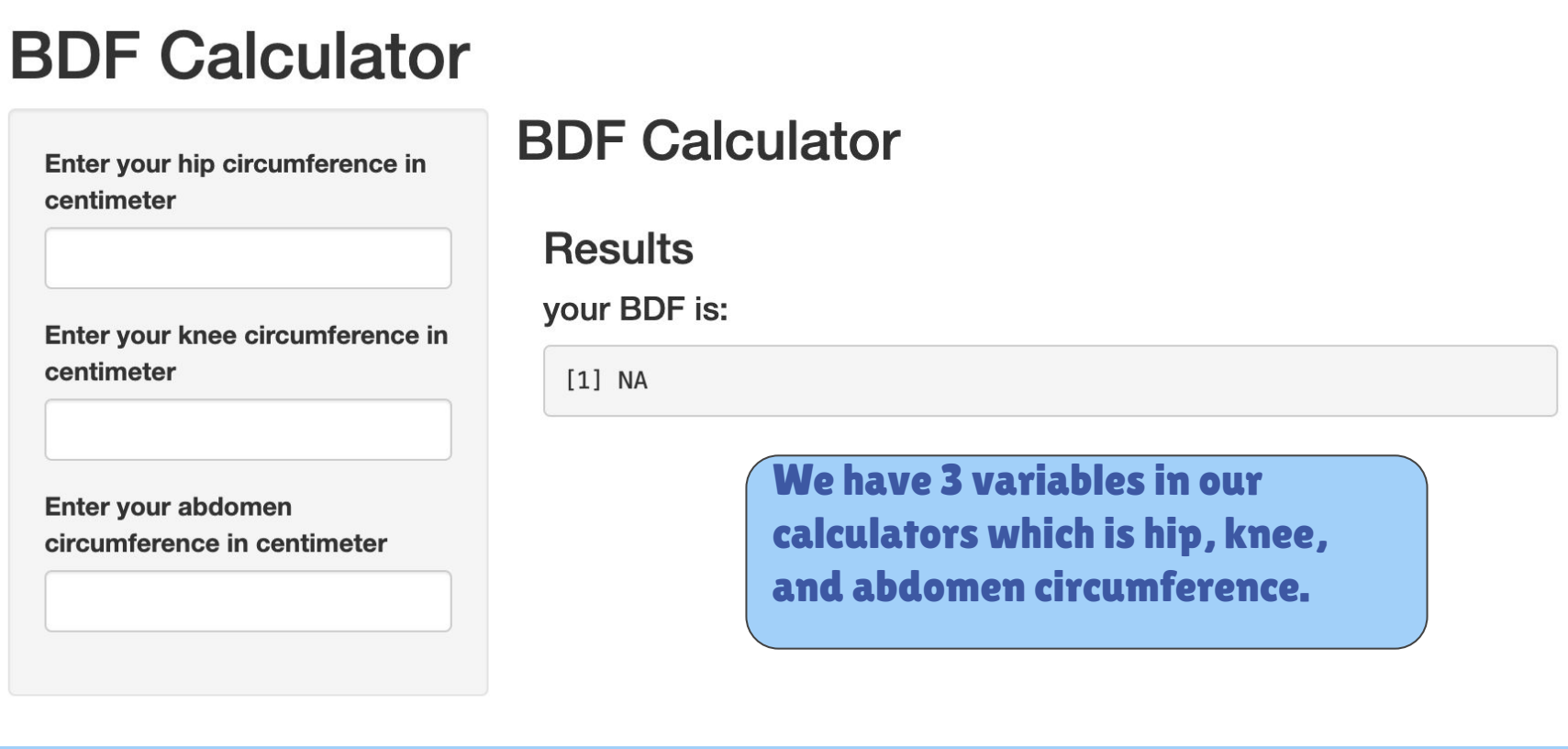


05

Results & Interpretation

BDF Calculator we made

http://127.0.0.1:7088/



We have 3 variables in our calculators which is hip, knee, and abdomen circumference.

Our model are more accurate compare with BMI

Real body fat: 12.3%

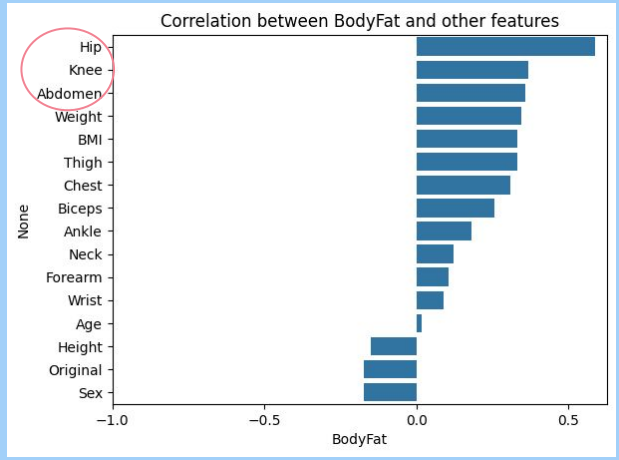
Our function: hip circumference: 94.5cm, knee circumference: 37.3, abdomen circumference: 85.2cm
Body fat: 16. 7308%

- **Most precise one**

BMI function: height: 172 cm, weight: 69.97 kg
Body fat: 23.7%

RMF function: RMF
Body fat: 35.3%

How do I made the BDF calculator?



3 numeric input

- Hip circumference in centimeters
- Knee circumference in centimeters
- Abdomen circumference in centimeters

Title Panel: “BDF Calculator”

BDF function: (-41.28+(0.93*hip)-(0.89*knee)+(0.039*abdomen))
We use 3 most related variables with body fat(hip, knee, and abdomen) to from a function in linear regression model.

Function forming

X = df (hip, knee, abdomen)
y = df (BodyFat)
From sklearn.linear model import LinearRegression

- We first create a linear regression model
- Then this model will train to features (X) and target (y)
- Get the coefficients and intercept
- Get the function

Limitation & Challenge

Linear regression model is a easy way to form equation using the relationship between multiple predictor variables and predicted. But when the datasets is large using linear regression might not be the best choice.

However, linear regression model can be affected by outlier, and inability to determine the feature importance in case of high multicollinearity.

It is not precise as Random Forest model. However, Random Forest model is not able to form a simple function, it is more complex and not easy to use in daily life.

06

Conclusion & Discussion

Main Takeaways

- After visualizing the data, we identified three body traits–**hip, knee, and abdomen circumference**–that are most correlated with body fat percentage.

- Our study used **multiple regression models** to predict body fat percentage based on **three body traits** from the dataset. Among the models tested, **Random Forest Regression** provided the most precise predictions. However, It's too complicated to analyze and use. Finally, we choose **Linear Regression.**

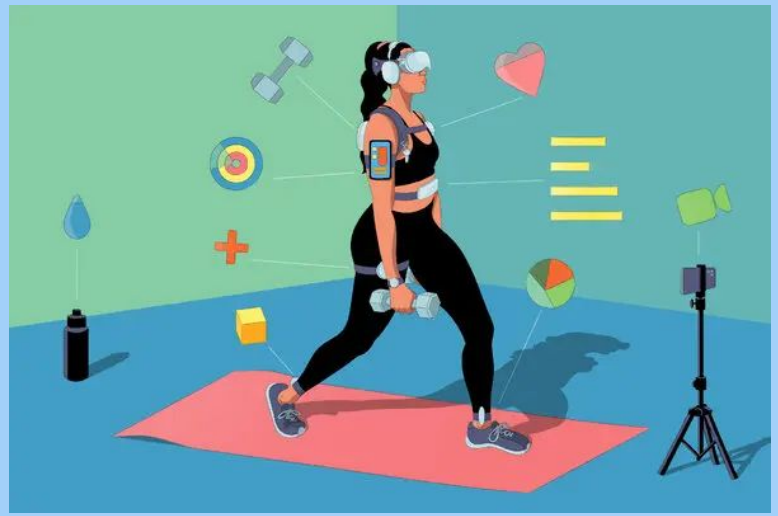
- Finally, we developed our **Body Fat Calculator**, which uses hip, knee, and abdomen circumference to estimate body fat percentage. Our calculator is **more accurate than BMI and RFM** in predicting body fat.

Implications & Applications

The findings also suggest that **BMI is not a reliable predictor** of body fat percentage, as it often misclassified individuals.

Our model, based on **hip, knee, and abdomen circumference**, offers a more **accurate and accessible method** for estimating body fat.

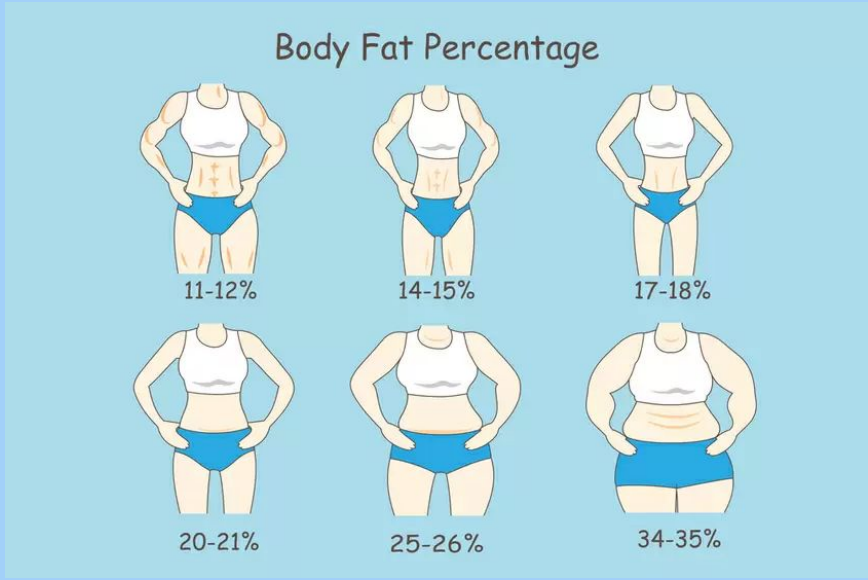
This can be used in **health assessments, fitness tracking.**The Body Fat Calculator also can be used in some mobile health apps.



Improvement and future direction

In the future, people can use our function to estimate their body fat online using the datasets they measured by themselves. This can help individuals better understand their body composition and make healthier decisions. It is a useful health tracker.

It could be used in an exercise or diet tracking app. Therefore, people would know what is affecting their body fat and how to gain more muscle and become healthier. It could be an effective tool related to our health.



07

References & Acknowledges

Citations for Data Sources & Paper References

World Health Organization - Global Health Observatory (2024) – processed by Our World in Data. “Obesity in adults” [dataset]. World Health Organization, “Global Health Observatory” [original data]. Retrieved March 11, 2025 from https://ourworldindata.org/grapher/share-of-adults-defined-as-obese

https://www.kaggle.com/datasets/simonezappatini/body-fat-extended-dataset

https://www.kaggle.com/code/elvinrustam/bodyfat-prediction-regression-tutorial

Woolcott, O.O., Bergman, R.N. Relative fat mass (RFM) as a new estimator of whole-body fat percentage — A cross-sectional study in American adult individuals. *Sci Rep* 8, 10980 (2018). https://doi.org/10.1038/s41598-018-29362-1

Woolcott, O. O., & Seuring, T. (2023). Temporal trends in obesity defined by the relative fat mass (RFM) index among adults in the United States from 1999 to 2020: a population-based study. *BMJ open*, *13*(8), e071295. https://doi.org/10.1136/bmjopen-2022-071295

Thanks!

Q & A Time!