Predictive Modeling in Strength Training Using Machine Learning

Kudakwashe Mavis Chakanyuka

23 March 2025

Research Goal

Objective :

- To build a predictive model for strength performance (1RM - 1 Repetition Maximum)

Focus :

- Squat, Bench Press, Deadlift

Why it matters?

- Personalized training recommendations
- Minimize injury risks
- Data-driven performance tracking

i) Squat

i) Bench Press

iii) Dead

Datasets Used

```
ar/folders/k_/xydd2f0x2cqgb7pwj5wc924m0000gn/T/ipykernel_3743/328005493.py:10: DtypeWarning: Columns (31,33,
df2 = pd.read_csv("/Users/kudakwashechakanyuka/Downloads/Queen Kuda Project/archive/openpowerlifting-2024-01
ar/folders/k_/xydd2f0x2cqgb7pwj5wc924m0000gn/T/ipykernel_3743/328005493.py:11: DtypeWarning: Columns (35) ha
df3 = pd.read_csv("/Users/kudakwashechakanyuka/Downloads/Queen Kuda Project/archive/openpowerlifting.csv")
  Age  BodyweightKg  Best3SquatKg  Best3BenchKg  Best3DeadliftKg  HeightCm  \
 35.0         137.1         327.5         235.0            365.0       175
 31.0          98.8         325.0         185.0            335.0       175
 27.0         132.0         412.5         237.5            427.5       162
 26.0         109.6         362.5         207.5            365.0       175
 21.0          99.5         340.0         245.0            340.0       175

       BMI  Sex_F  Sex_M
 44.767347  False   True
 32.261224  False   True
 50.297211   True  False
 35.787755  False   True
 32.489796  False   True
```

Three Data Sources :

* **Powerlifting Rankings Dataset**
  Elite lifters' competition data
* **Kaggle Strength Dataset**
  Gym-goers of all levels
* **General Fitness Metrics**
  Beginners & casual lifters

Key Features Selected

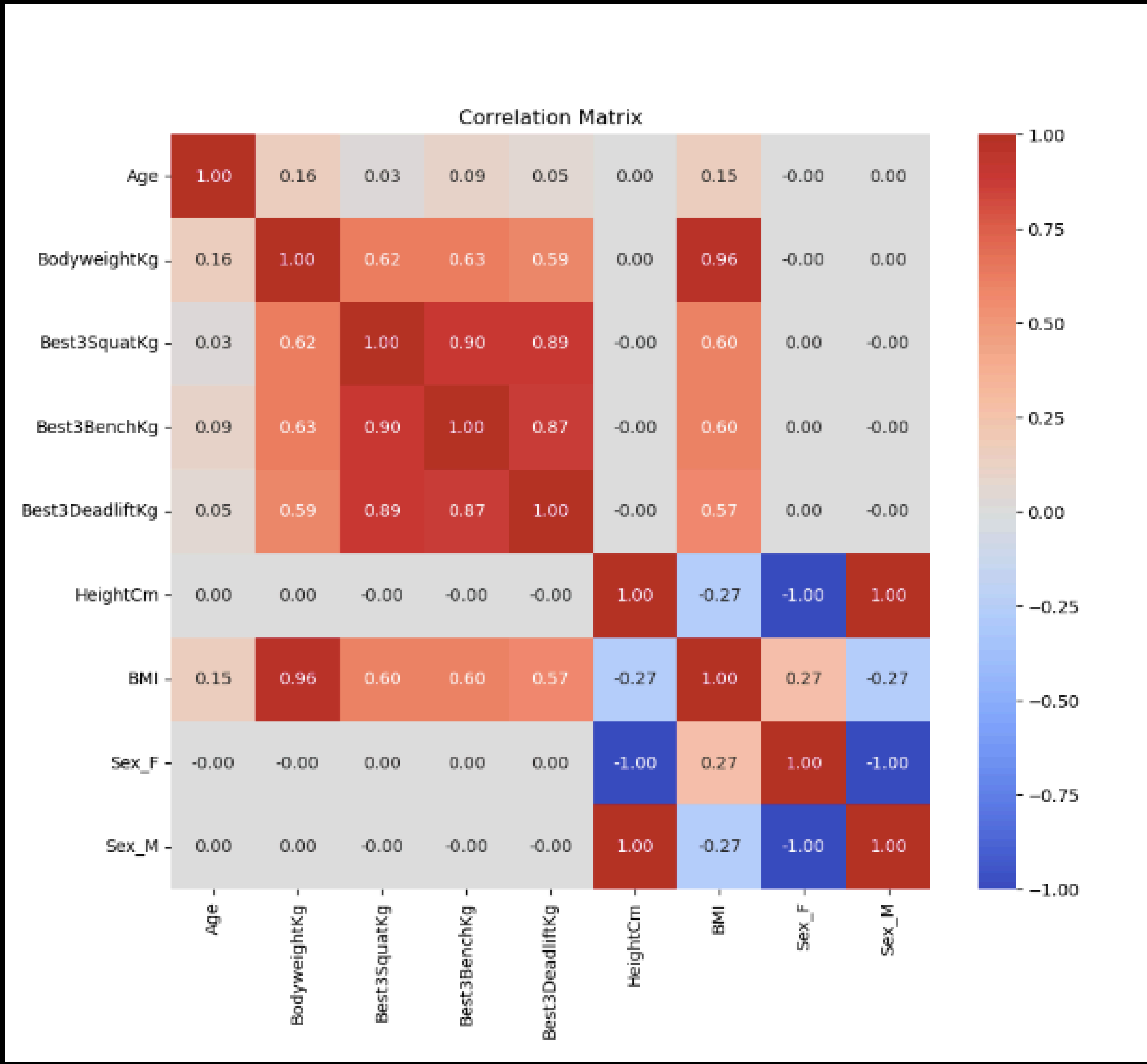
Age, Gender, Body Weight, Training History, Performance Metrics (1RM

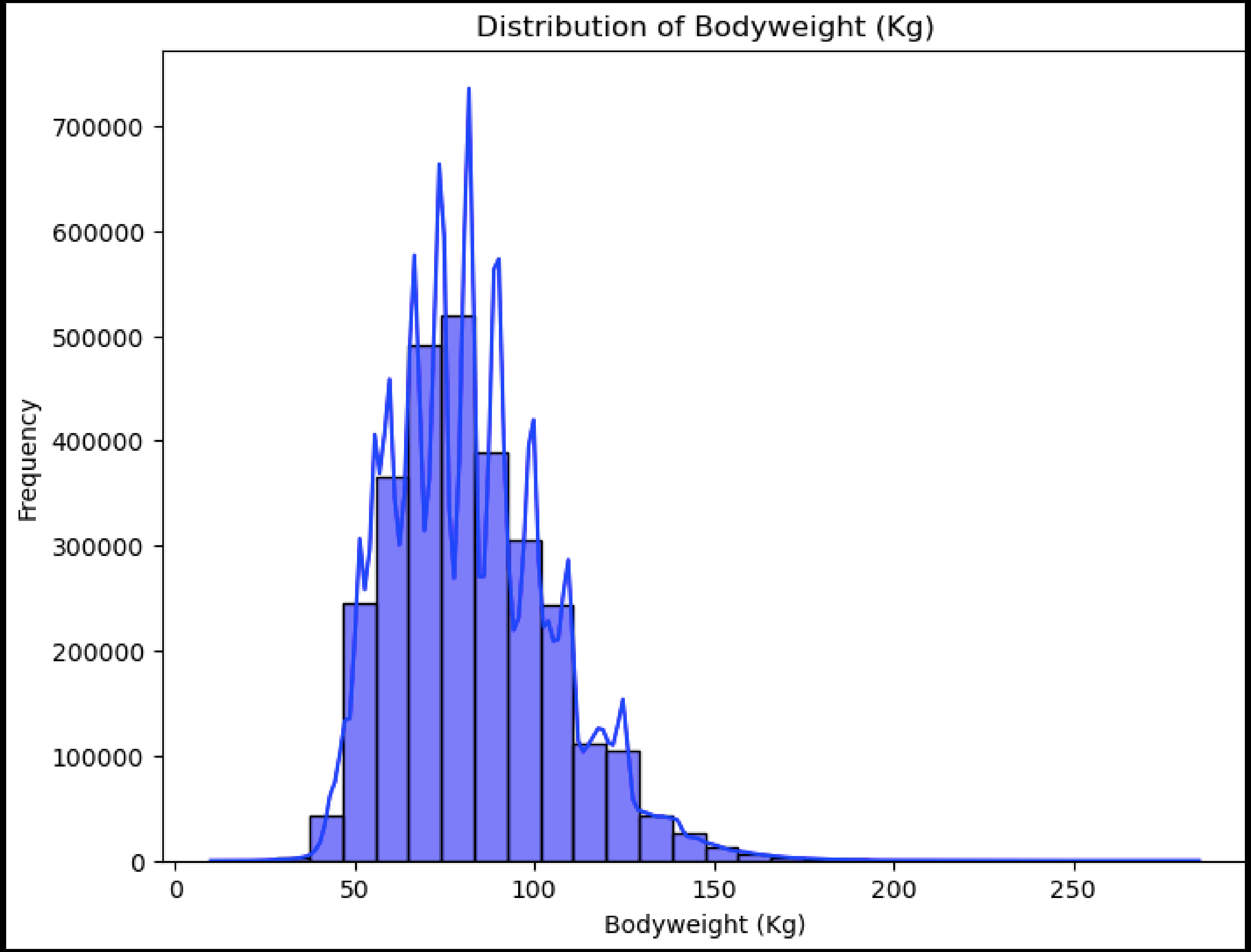
Data Processing & Feature Selection Preprocessing Steps

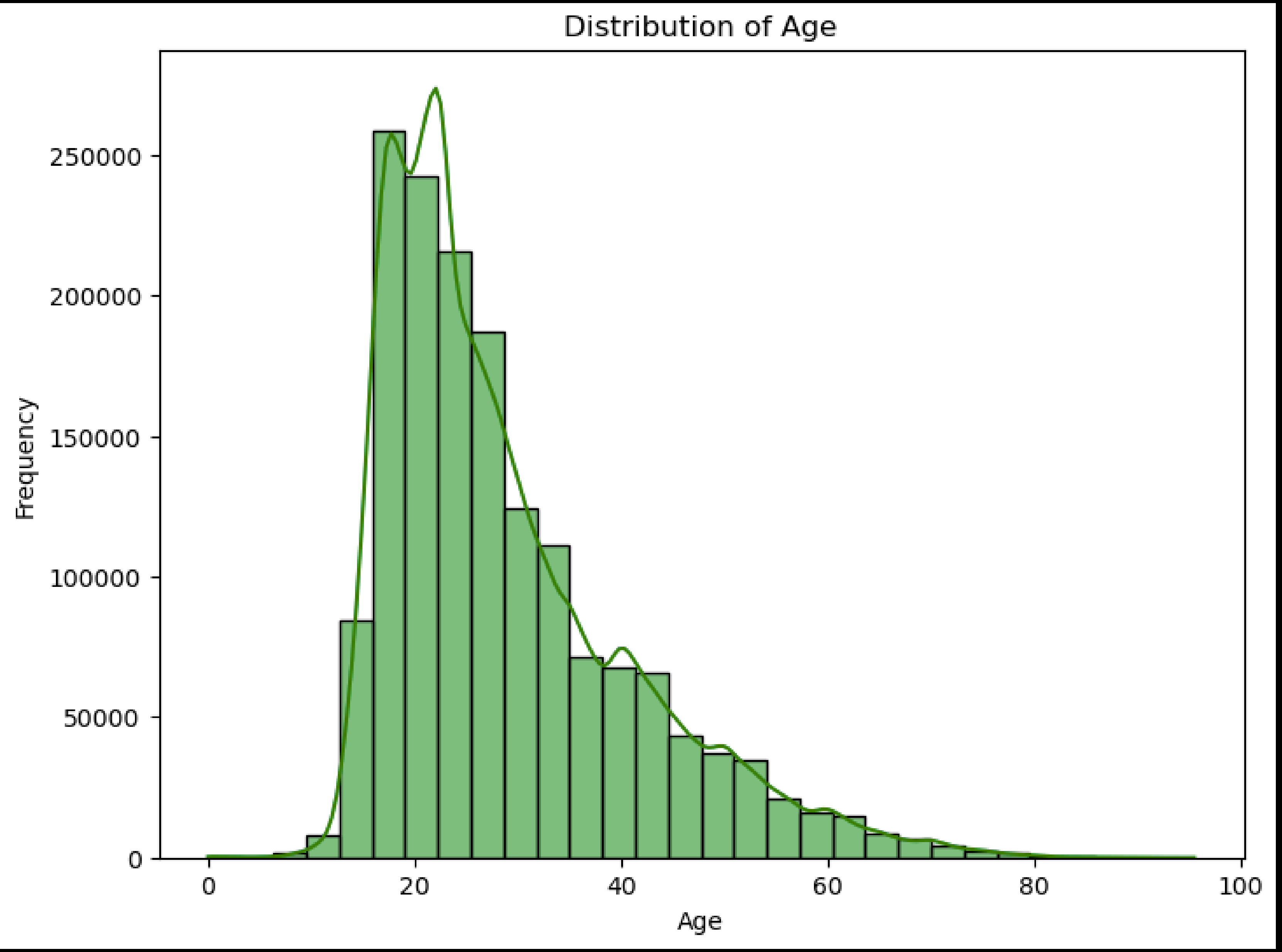
- Standardized common features (age, weight, etc.)
- Handled missing values (imputation)
- Encoded categorical data
- Removed extreme outliers

Key Variables Retained:

- Demographics: Age, gender, weight, height, BMI
- Training History: Frequency, experience, workout intensity
- Performance Data: Past 1RM values, sets, reps, RPE








```
...   Squat - Training Set Shape (X_train_squat): (2347420, 5)
      Squat - Testing Set Shape (X_test_squat): (586855, 5)
      Bench Press - Training Set Shape (X_train_bench): (2347420, 5)
      Bench Press - Testing Set Shape (X_test_bench): (586855, 5)
      Deadlift - Training Set Shape (X_train_deadlift): (2347420, 5)
      Deadlift - Testing Set Shape (X_test_deadlift): (586855, 5)
```

Machine Learning Models Used

Target Variable :

- 1RM (One-Rep Max) for each lift

Models Applied :

- Decision Trees – Simple, interpretable baseline
- Bagging – Reduces overfitting, improves stability
- Random Forest – Best model for capturing feature interactions

Why Random Forest?

- Handles non-linear relationships well
- Works with diverse datasets
- Provides reliable and robust predictions

```
RandomForestRegressor
RandomForestRegressor(n_estimators=50, n_jobs=-1, random_state=42)
```

Model Training & Evaluation

```
Skipping R² calculation for Squat Model due to insufficient test samples.
Evaluation Metrics for Squat Model:
Mean Absolute Error (MAE): 14.19
Mean Squared Error (MSE): 464.93
Root Mean Squared Error (RMSE): 21.56
R-squared (R²): Undefined (not enough samples)

Skipping R² calculation for Bench Press Model due to insufficient test samples.
Evaluation Metrics for Bench Press Model:
Mean Absolute Error (MAE): 14.19
Mean Squared Error (MSE): 464.93
Root Mean Squared Error (RMSE): 21.56
R-squared (R²): Undefined (not enough samples)

Skipping R² calculation for Deadlift Model due to insufficient test samples.
Evaluation Metrics for Deadlift Model:
Mean Absolute Error (MAE): 14.19
Mean Squared Error (MSE): 464.93
Root Mean Squared Error (RMSE): 21.56
R-squared (R²): Undefined (not enough samples)
```

Separate Models for Each Lift :

Squat, Bench Press, Deadlift trained independently

Cross-Validation Applied – Ensured generalization to unseen data

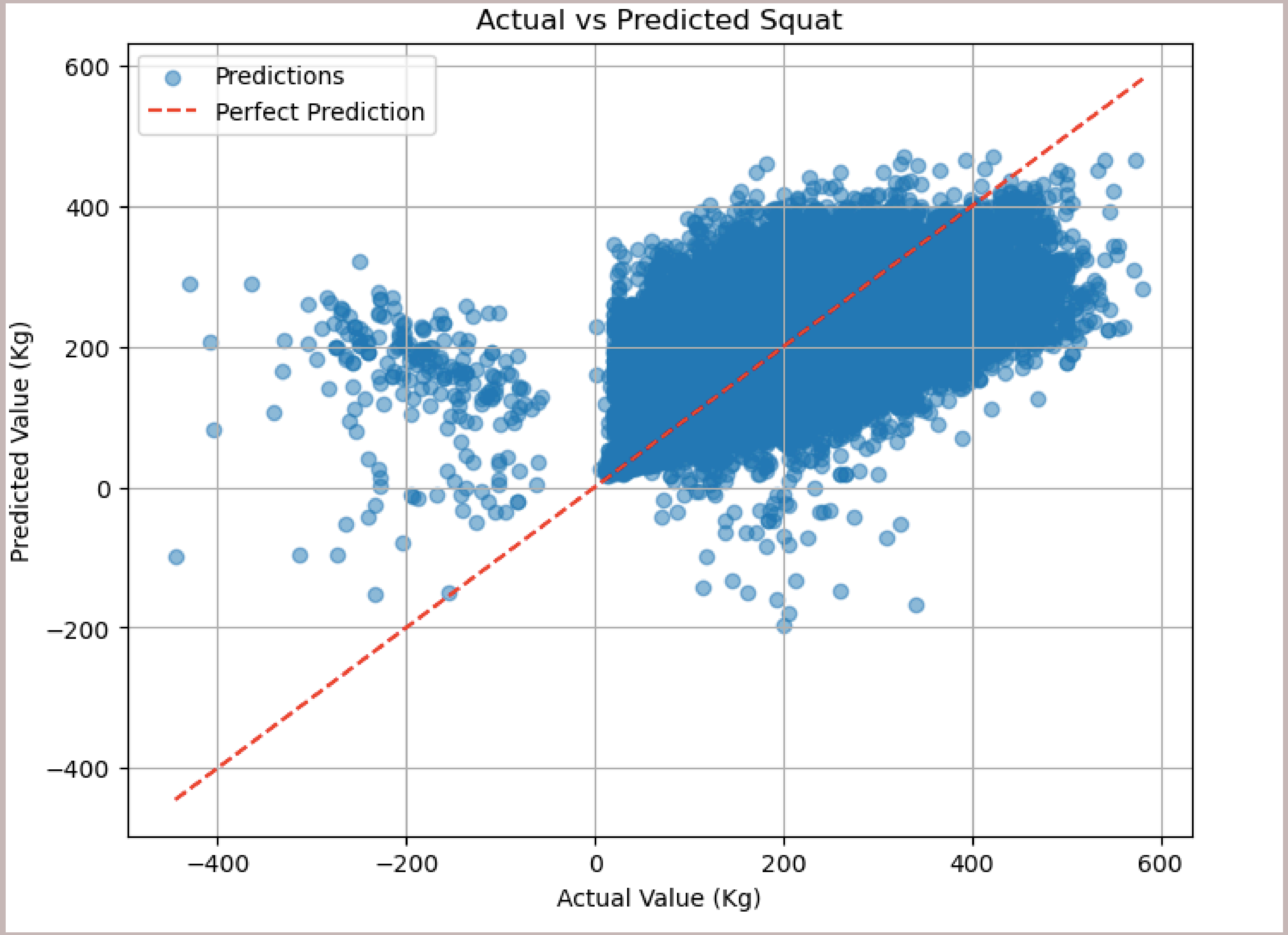
Evaluation Metrics Used :

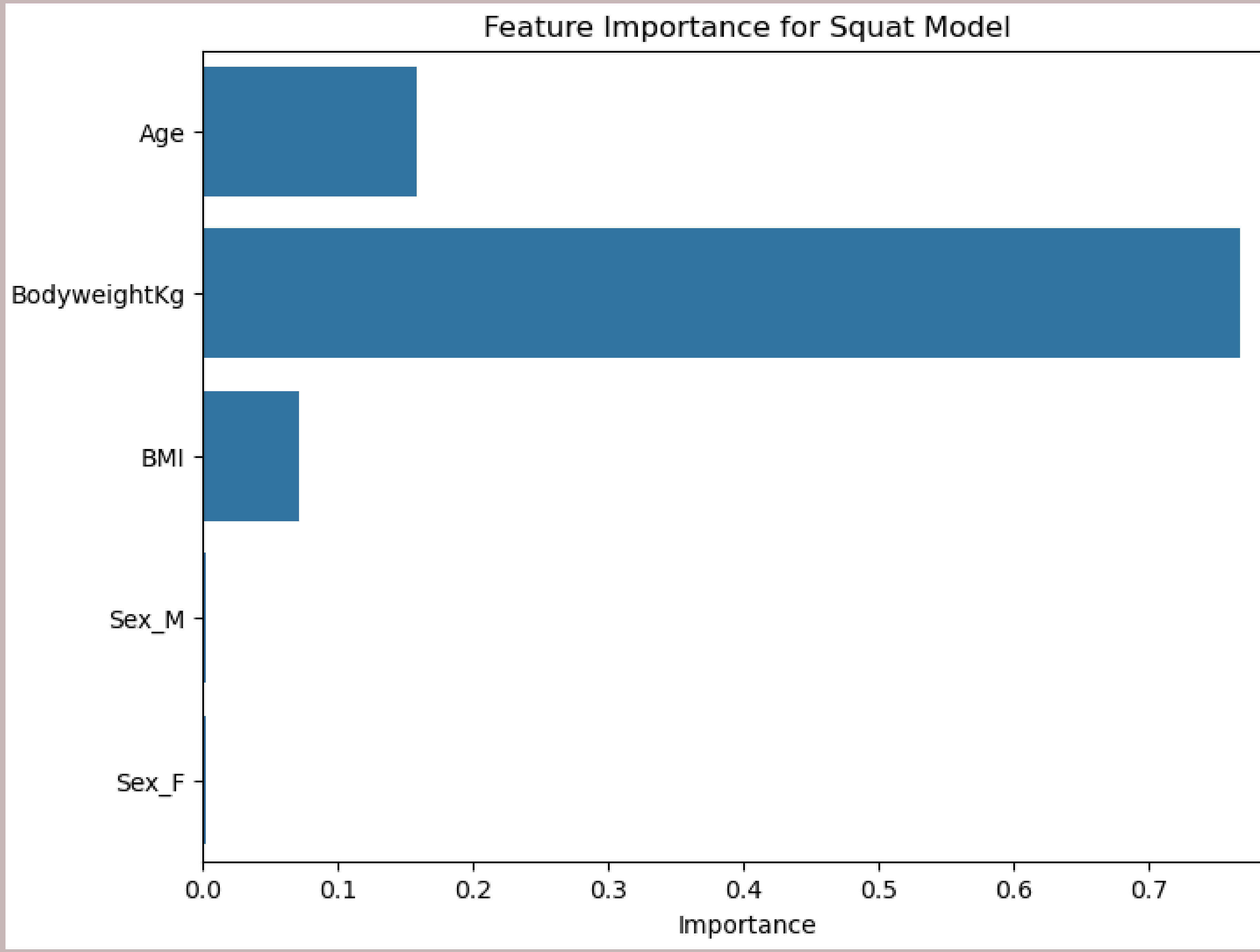
MAE (Mean Absolute Error): Measures avg. error

RMSE (Root Mean Squared Error): Penalizes large errors

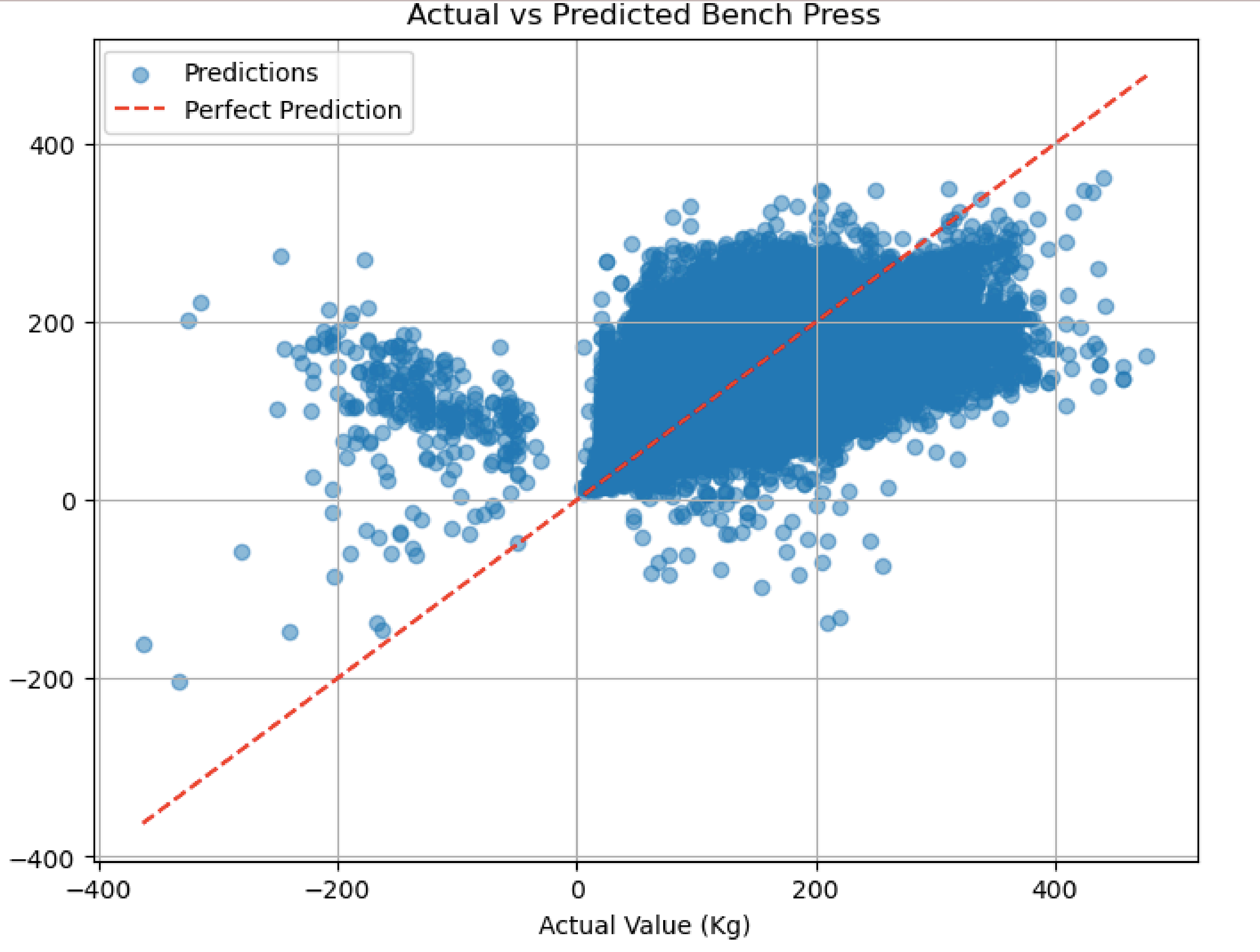
R^2 Score: Measures how well the model explains variance

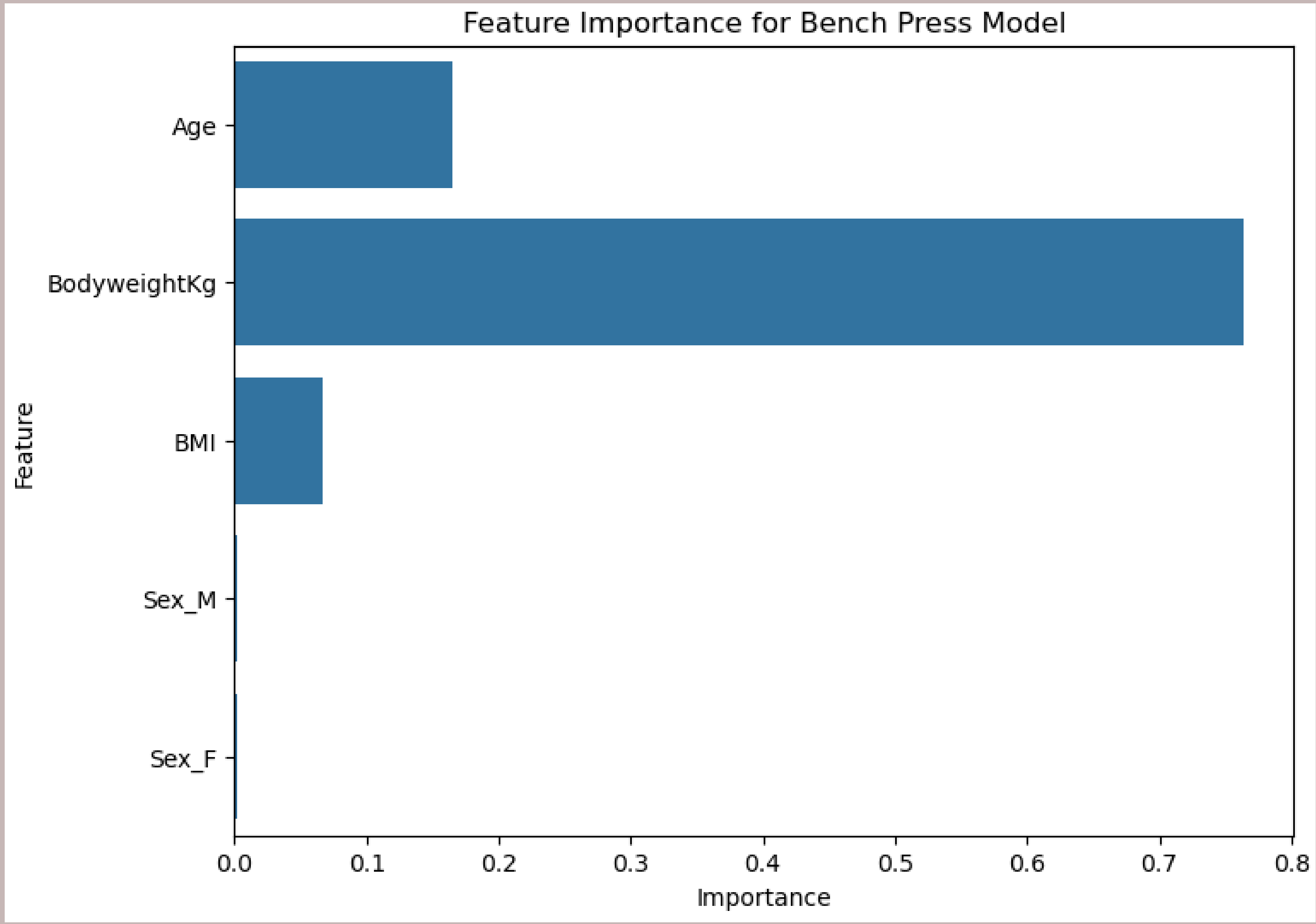
Findings from Model



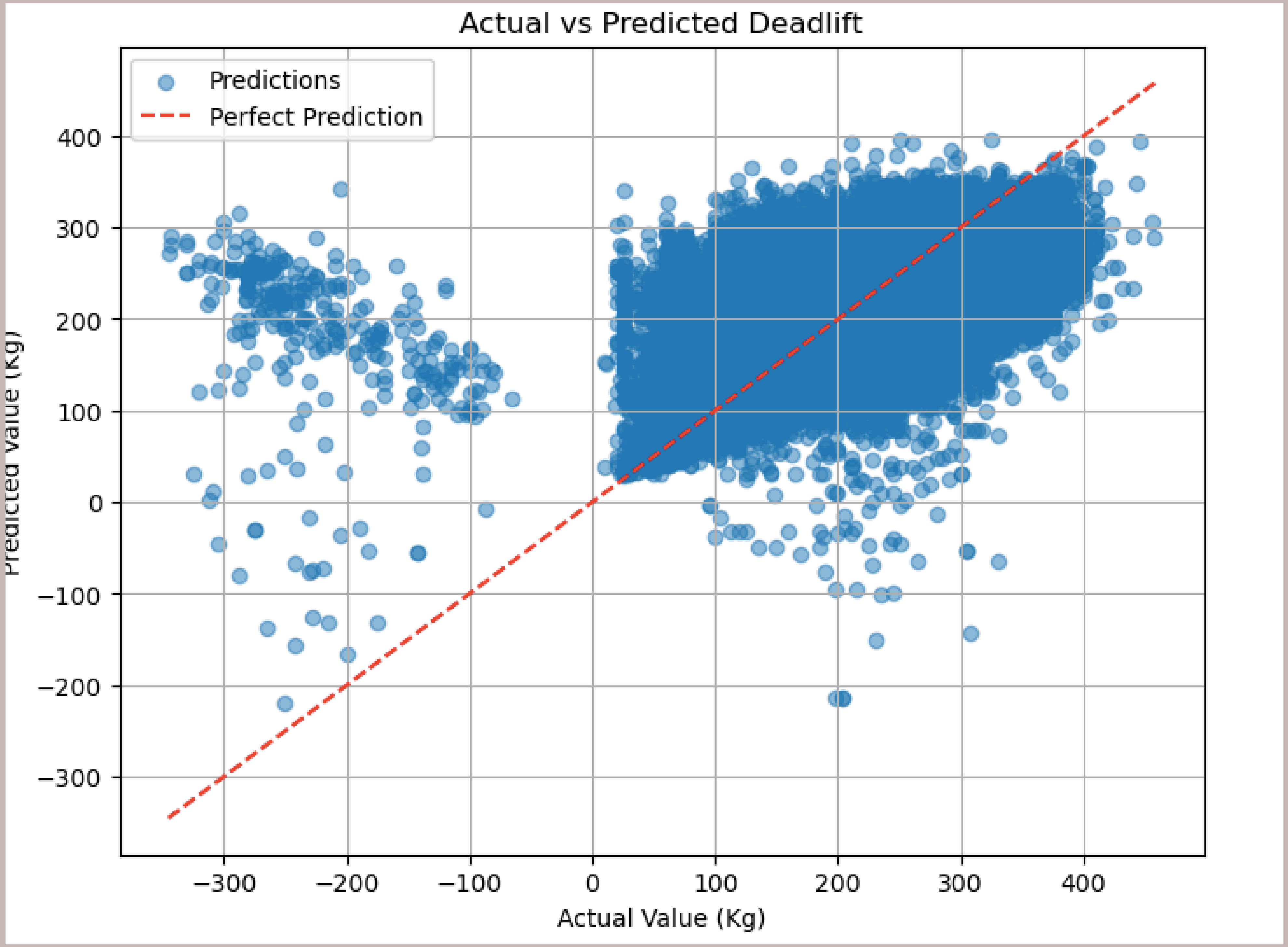


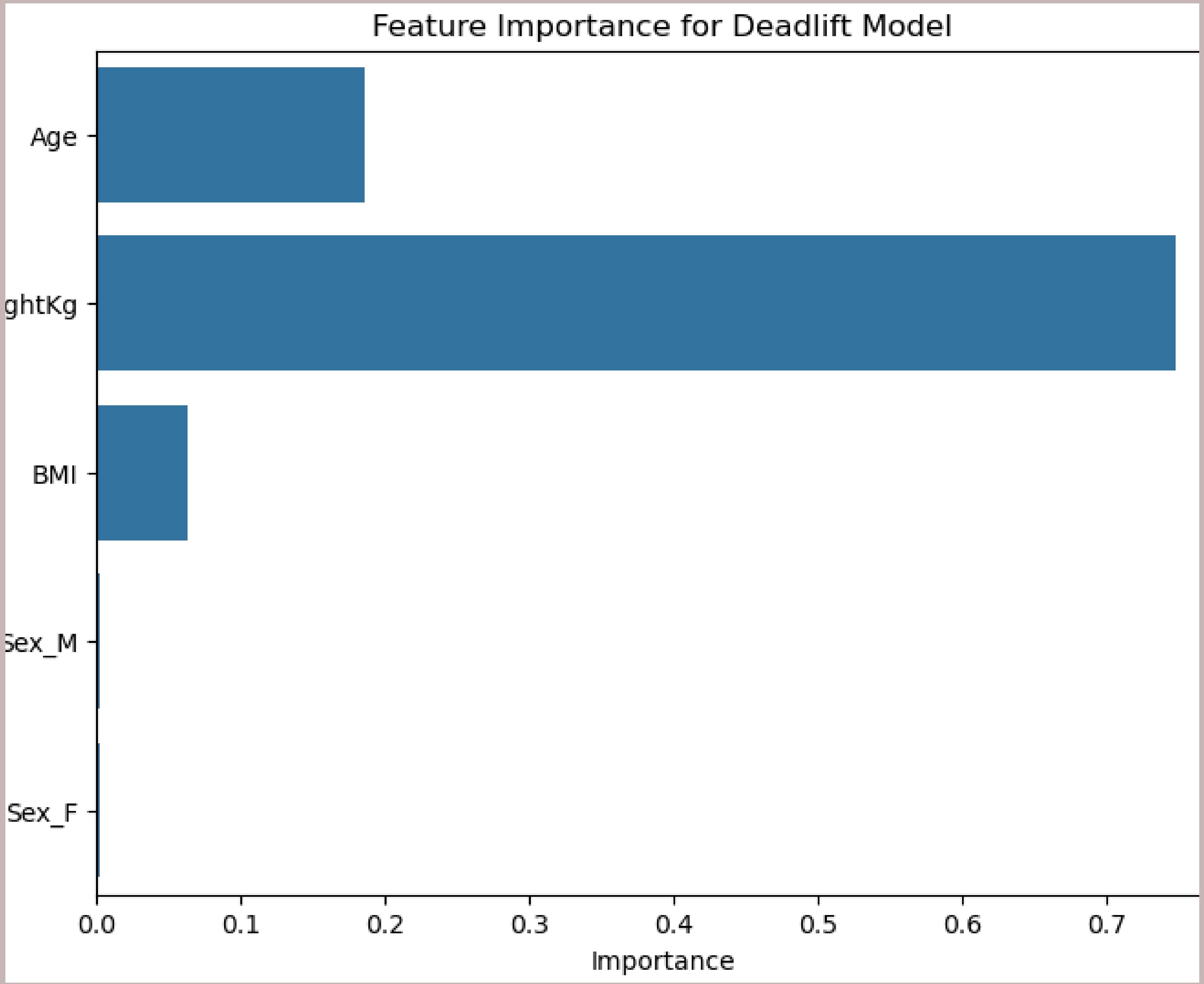
Findings from Model continued...





Findings from Model continued....





Results Analysis

Squat :

- Strong Correlation: The data revealed a strong positive correlation between body weight and squat 1RM (one-rep max), emphasizing the significant role body mass plays in generating lower-body strength.

Bench Press :

- Significant Predictor: The analysis has revealed that training consistency and experience level were top predictors of bench press 1RM (one-rep max), demonstrating the crucial role of structured routines and progressive upper-body strength development.

Deadlift :

- Strong Correlation: The analysis demonstrated a strong positive correlation between posterior chain strength and deadlift 1RM (one-rep max), highlighting the critical role of hamstrings, glutes, and lower back in achieving optimal lifting capacity.

Strength Multipliers & Practical Application

Exercise	Beginner Multiplier	Advanced Multiplier
Squat	0.75	1.25
Bench Press	0.80	1.20
Deadlift	0.70	1.30

How These Work :

- Predicts expected strength based on body weight
- Helps users set realistic training goals
- Used for progress tracking

User Interaction Model

User Inputs :

- Age, Gender, Body Weight, Training Frequency

Model Output :

- Predicted 1RM values for Squat, Bench, Deadlift

Application :

- Gym-goers & trainers can use predictions for personalized training plans

Key Observations :

- Body weight is the strongest predictor for squat & deadlift
- Training frequency & consistency impact bench press performance
- Beginners progress at different rates based on prior athletic experience

Model Performance :

- Random Forest consistently outperformed Decision Trees
- MAE and RMSE values confirmed model accuracy

Age 50

Bodyweigh... 92

Height (cm) 171

Sex F

Disability No

Experience... B

Predict 1RM

```
Input Data for Prediction:
   age  bodyweightkg  best3squatkg  best3benchkg  best3deadliftkg  heightcm  \
0   50          92.0             0             0                0     171.0

         bmi  sex_f  sex_m
0  31.462672      1      0

Predicted 1RM Values:
Squat 1RM: 27.6 kg
Bench Press 1RM: 27.6 kg
Deadlift 1RM: 27.6 kg
```

Age 50

Bodyweigh... 92

Height (cm) 171

Sex M

Disability No

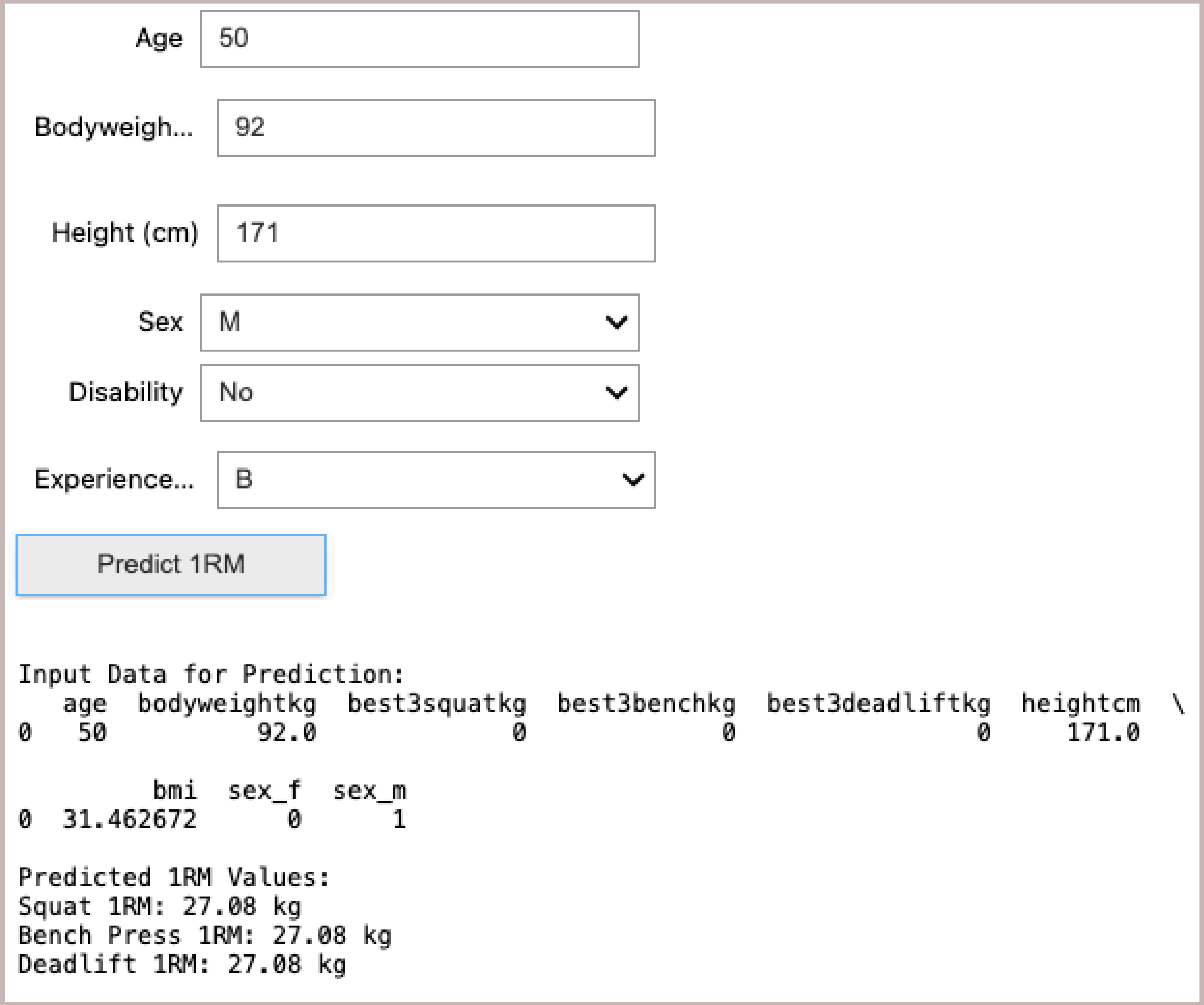
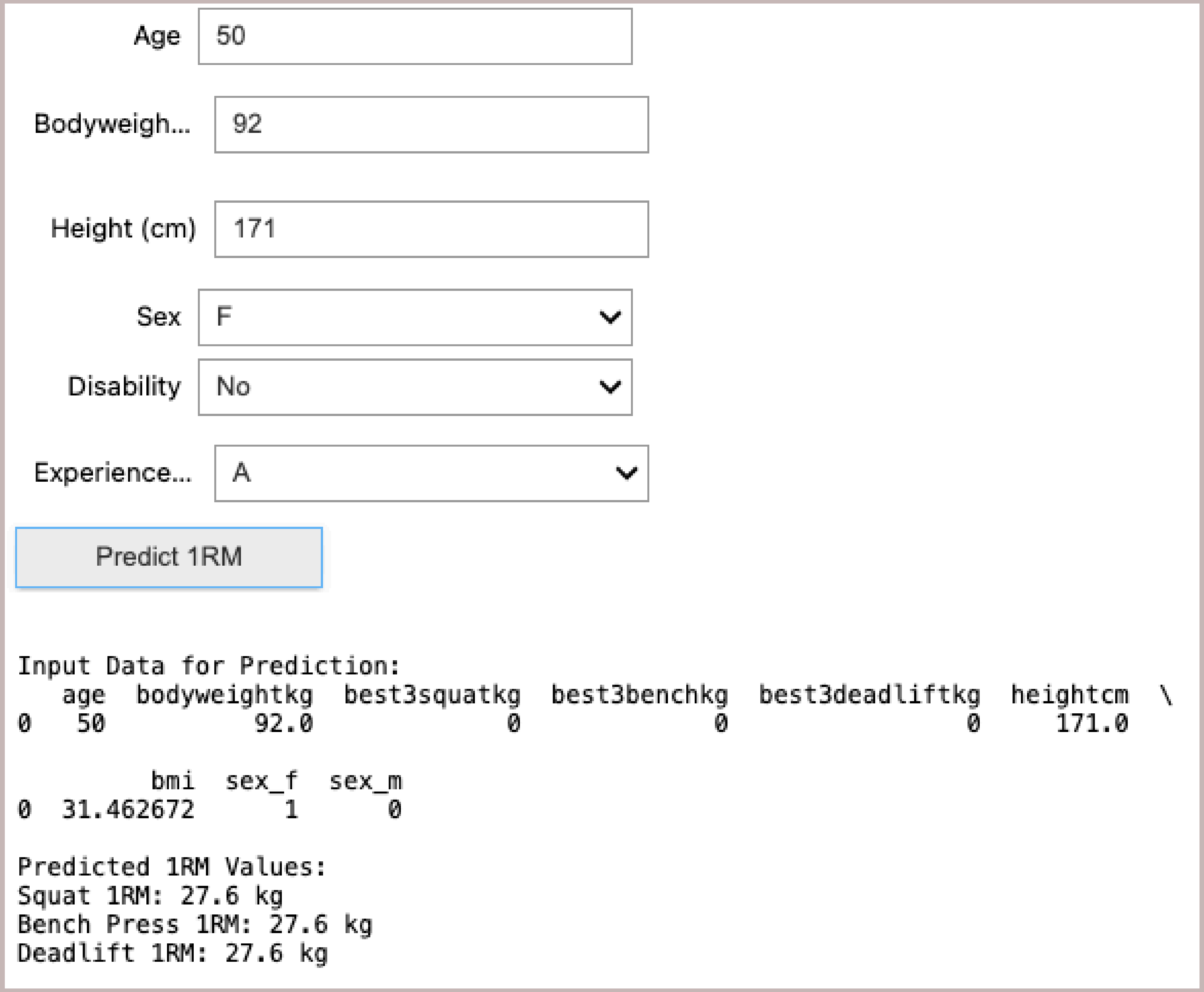
Experience... A

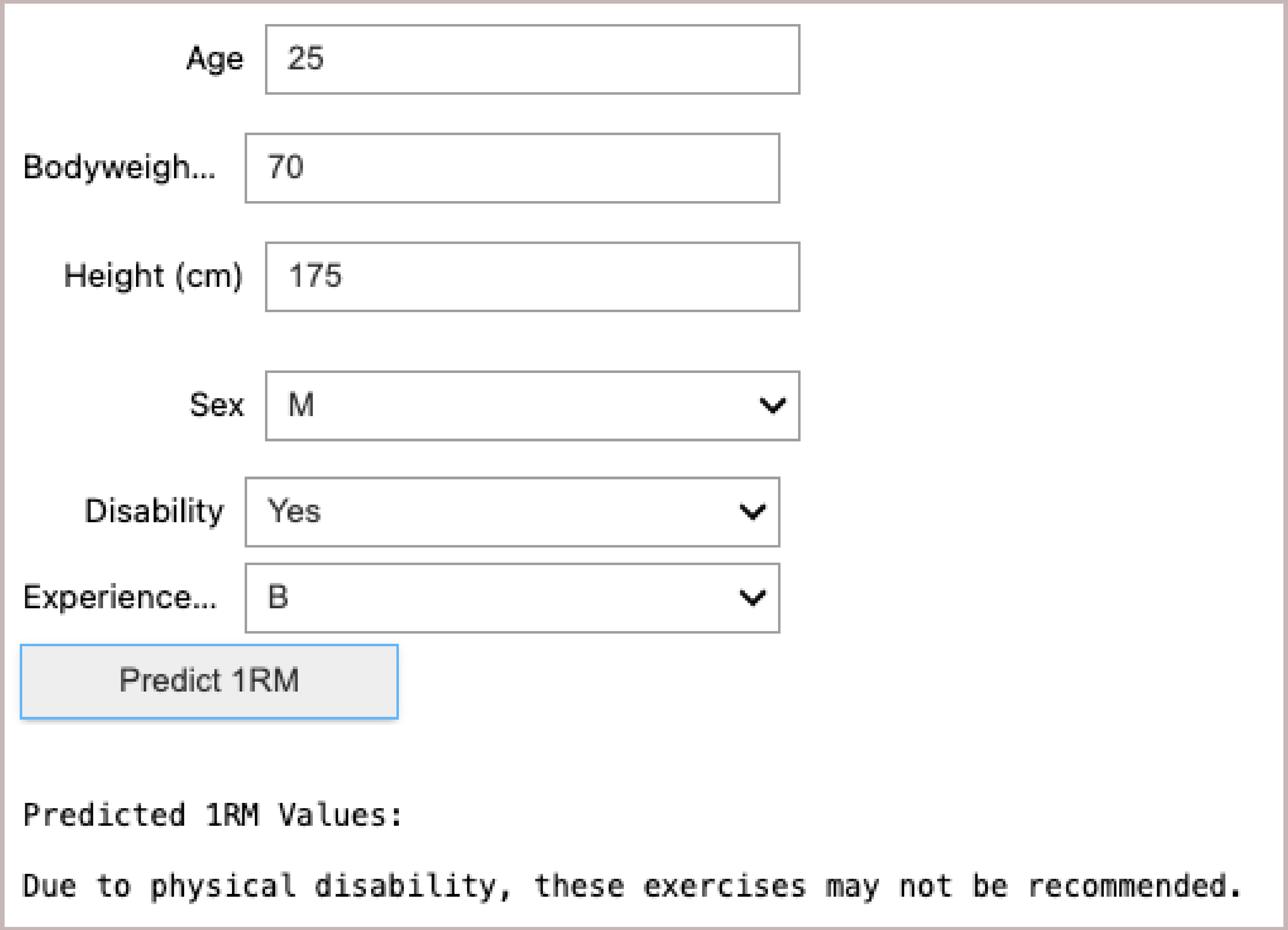
Predict 1RM

```
Input Data for Prediction:
   age  bodyweightkg  best3squatkg  best3benchkg  best3deadliftkg  heightcm  \
0   50          92.0             0             0                0     171.0

         bmi  sex_f  sex_m
0  31.462672      0      1

Predicted 1RM Values:
Squat 1RM: 27.08 kg
Bench Press 1RM: 27.08 kg
Deadlift 1RM: 27.08 kg
```


Age 50
Bodyweigh... 92
Height (cm) 171
Sex M
Disability No
Experience... B
Predict 1RM
Input Data for Prediction:
age bodyweightkg best3squatkg best3benchkg best3deadliftkg heightcm \
0 50 92.0 0 0 0 171.0
bmi sex_f sex_m
0 31.462672 0 1
Predicted 1RM Values:
Squat 1RM: 27.08 kg
Bench Press 1RM: 27.08 kg
Deadlift 1RM: 27.08 kg

Age 50
Bodyweigh... 92
Height (cm) 171
Sex F
Disability No
Experience... A
Predict 1RM
Input Data for Prediction:
age bodyweightkg best3squatkg best3benchkg best3deadliftkg heightcm \
0 50 92.0 0 0 0 171.0
bmi sex_f sex_m
0 31.462672 1 0
Predicted 1RM Values:
Squat 1RM: 27.6 kg
Bench Press 1RM: 27.6 kg
Deadlift 1RM: 27.6 kg


Age
25
Bodyweigh...
70
Height (cm)
175
Sex
M
Disability
Yes
Experience...
B
Predict 1RM
Predicted 1RM Values:
Due to physical disability, these exercises may not be recommended.

Key Takeaways

- Machine learning techniques demonstrate strong effectiveness in predicting strength performance, showcasing their potential in fitness analytics.
- Optimal Model : Random Forest Regression emerged as the most accurate method, effectively handling complex relationships between variables.
- Strength Multipliers : Provides practical benchmarks for lifters, aiding in goal-setting and performance evaluation.

Impact on Strength Training

- Personalized Training Plans : Tailored recommendations based on individual attributes and predicted performance.
- Data-Driven Tracking : Enables precise monitoring of progress and identification of areas for improvement.
- Injury Prevention : Reduces risk by suggesting appropriate training intensities and loads.

The End!