CALORIES BURNT PREDICTION

OBJECTIVE

The main objective of this machine learning project is to build a predictive model that accurately estimates the number of calories burnt by an individual during physical activity. This can be highly beneficial for fitness tracking applications, personalized training plans, and health monitoring systems. By inputting parameters such as age, gender, height, weight, heart rate, and body temperature during exercise, the model will predict the calorie expenditure of a person.

DATASET USED

We used **two datasets**:

1. `calories.csv`: This dataset contains the `User_ID` and corresponding `Calories` burnt by that user during their exercise session.
2. `exercise.csv`: This dataset contains detailed exercise-related features such as:
    - `User_ID`
    - `Gender`
    - `Age`
    - `Height`
    - `Weight`
    - `Duration` (minutes of activity)
    - `Heart_Rate`
    - `Body_Temp` (Body temperature in °C)

Both datasets were merged on `User_ID` to form a comprehensive dataset with input features and the target variable (`Calories`)

MODEL CHOSEN

We have chosen the XGBoost Regressor as the final and most suitable machine learning model for this project.

Why XGBoost?

Multiple regression models were trained and evaluated, including:

- XGBoost

- Linear Regression

- Random Forest Regressor

Model Performance Summary:

XGBoost | Accuracy: 98.34% | MAE: 1.48

Linear Regression | Accuracy: 90.62% | MAE: 8.39

Random Forest | Accuracy: 98.11% | MAE: 1.69

As seen in the results, **XGBoost outperformed the other models** by achieving the **highest accuracy (98.34%)** and the **lowest Mean Absolute Error (MAE) of 1.48**. While Random Forest also performed well, XGBoost offered a slightly better combination of accuracy and precision, and also tends to be more efficient and scalable for real-world datasets.

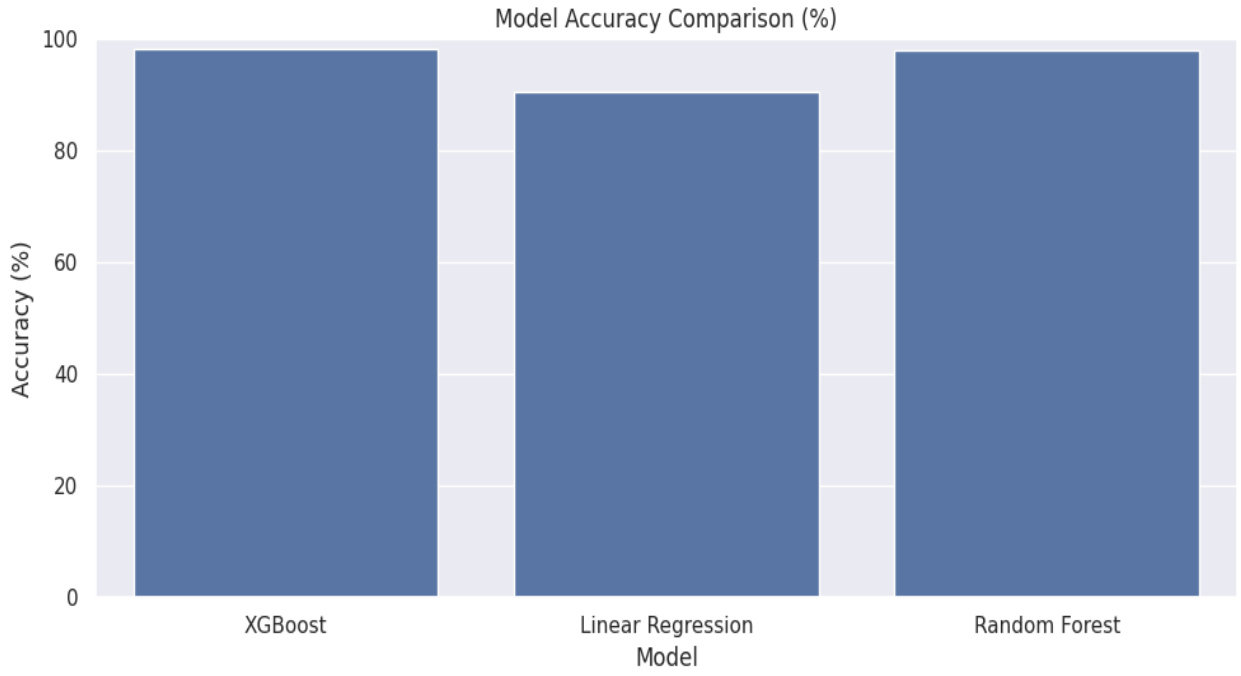
XGBoost's internal regularization and gradient boosting mechanism makes it resistant to overfitting and particularly well-suited for structured data like the one used in this project.

How was Accuracy Calculated?

Unlike classification problems, regression tasks don't have a standard definition for accuracy. However, we used a custom accuracy metric based on Mean Absolute Error (MAE):

accuracy = 100 - (MAE / mean_actual * 100)

This formula gives a percentage representing how close the predictions are to the actual average. Given its superior performance, XGBoost was selected as the final model.



PERFORMANCE METRICS

The model's performance was evaluated using the following regression evaluation metrics:

- Mean Absolute Error (MAE): Measures the average magnitude of errors in a set of predictions.
  → MAE = 1.48
  Interpretation: On average, the model's predictions are only 1.48 units off from the actual calorie values.
- Mean Squared Error (MSE): Measures the average of the squares of the errors.
  → MSE = 4.71
  Interpretation: Squared errors penalize larger errors more heavily, and our value is low, indicating good performance.

- Root Mean Squared Error (RMSE): The square root of MSE, which brings the error metric back to the original unit.
  → RMSE = 2.17
  Interpretation: The model has an average error of 2.17 calories in its predictions.
- R^2 Score (Coefficient of Determination): Measures how well the variation of the output is explained by the inputs.
  → R^2 Score = 0.9988
  Interpretation: 99.88% of the variance in calorie burn is explained by the features in our model, indicating an exceptionally accurate model.

CHALLENGES & LEARNINGS

One of the main challenges was ensuring the model did not overfit to the training data while still achieving high accuracy. We addressed this by comparing training and testing accuracies and carefully tuning the XGBoost model. Another learning experience was creating a meaningful accuracy metric for regression, as common classification metrics like precision and recall are not applicable. Additionally, saving the trained model and integrating it into a user-friendly interface using Streamlit helped reinforce real-world deployment skills.