Model Performance Evaluation Report

Project Title: Prediction of the Target Variable (Calories) Burnt During a Workout Session

Prepared by: Nene Chibuike Gideon

Objective:

The goal of this project is to analyze the provided dataset, perform Exploratory Data Analysis (EDA) to uncover insights, and build a linear regression model to predict the number of calories burnt during a workout session.

2. Performance Metrics

2.1 Mean Squared Error (MSE)

- **Train MSE: 0.145**
- **Test MSE: 0.143**

MSE quantifies the average squared difference between actual and predicted values. A lower MSE indicates better predictive accuracy. The close proximity of training and testing MSE values suggests that the model does not exhibit significant overfitting.

2.2 R^2 Score (Coefficient of Determination)

- **Train R^2: 0.9884**
- **Test R^2: 0.9889**

The R^2 score measures how well the independent variables explain the variance in the dependent variable. With an R^2 value of approximately **0.988**, the model explains **98.8%** of the variance in the target variable, indicating a strong predictive performance. The minimal difference between the training and testing R^2 scores suggests that the model generalizes well.

3. Model Evaluation

- The low MSE values indicate that the model's predictions are relatively accurate. However, the absolute value of MSE should be considered in the context of the dataset's scale.
- The high R^2 score implies that the model explains most of the variance, signifying strong predictive power.
- The small gap between training and testing metrics suggests minimal overfitting, indicating that the model generalizes well to unseen data.

4. Recommendations for Improvement

Although the model performs well, further refinements can be considered to enhance performance:

4.1 Hyperparameter Tuning

- Optimize hyperparameters using techniques such as Grid Search or Randomized Search.
- If using a linear regression model, explore polynomial features or regularization techniques (Lasso, Ridge).
- For tree-based models, adjust the depth, learning rate, or number of estimators.

4.2 Model Selection

- If further improvement is needed, consider ensemble methods (e.g., Random Forest, Gradient Boosting) or deep learning models for complex datasets.
- Evaluate simpler models to ensure the current complexity is justified.

5. Conclusion

The model exhibits strong performance with an R^2 score of **0.988** and a relatively low MSE, indicating high predictive accuracy. Given the minimal difference between training and testing performance, the model effectively generalizes to unseen data. While the model is already well-optimized, applying hyperparameter could further enhance its accuracy.