Calorie Predictor Report

# 1. Features Used

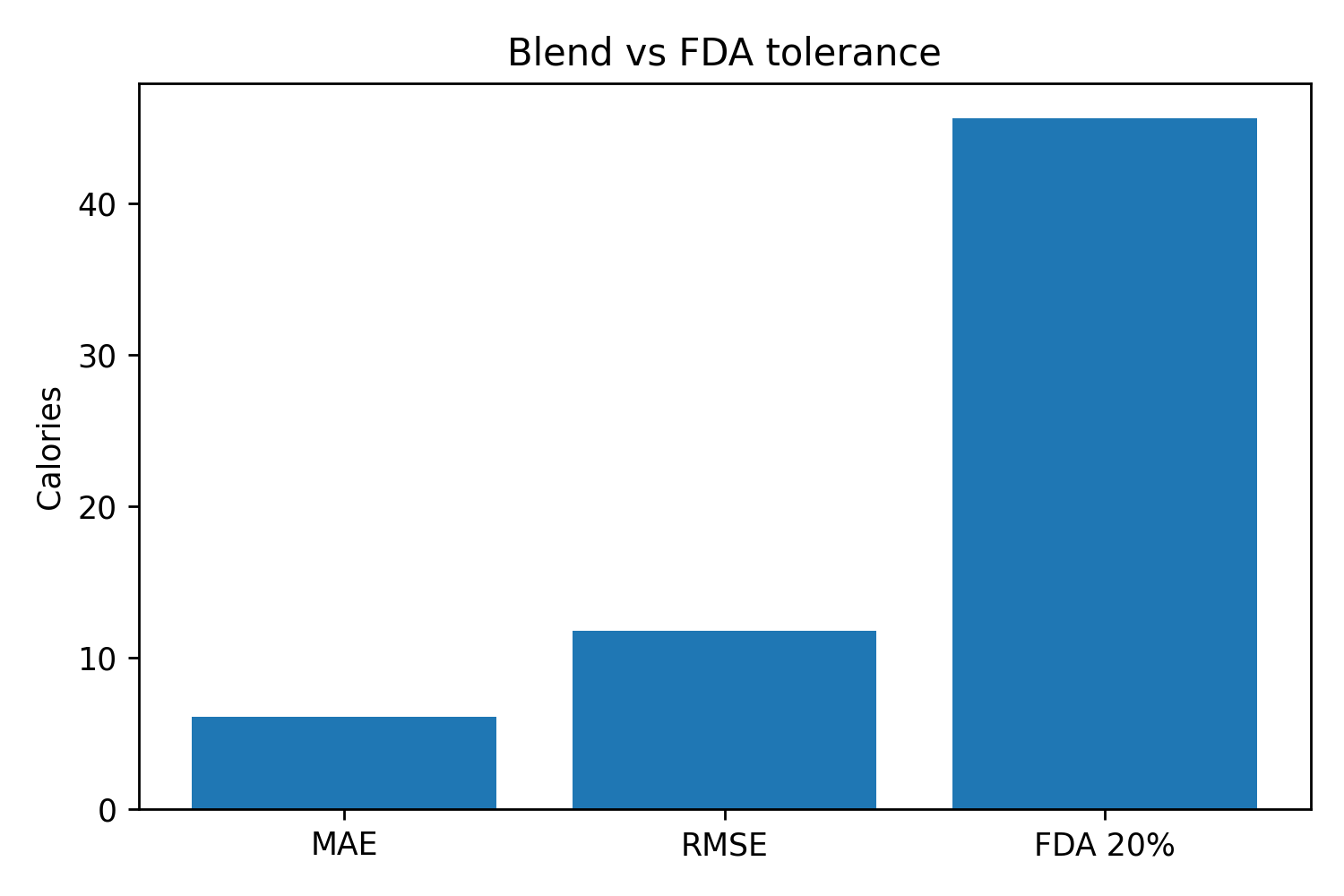
Mandatory: protein, carbohydrate, total\_fat, serving\_weight, calories

Optional: saturated\_fat, fiber, sugar, sodium

# 2. Results

|  |  |  |  |
| --- | --- | --- | --- |
| Model | RMSE | MAE | R² |
| HGB | 13.80 | 7.36 | 0.993 |
| Ridge | 12.12 | 6.19 | 0.995 |
| Blend | 11.78 | 6.10 | 0.995 |

# 3. FDA Comparison

The following graph shows the comparison between model metrics and FDA's 20% calorie-label tolerance.

**Why Not Use Multilinear Regression?**

While **multilinear regression** is a valid model for predicting calorie intake, there are a few reasons why we might prefer models like **HistGradientBoosting** and **Ridge Regression** in this scenario:

1. **Multilinear Regression Assumptions**: Linear regression assumes that the relationships between the features and the target variable (in this case, calories) are linear. If the relationships between your features (like protein, carbohydrates, fat, etc.) and calories are non-linear, the model might not capture them effectively. **HistGradientBoosting** and **Ridge Regression** can handle more complex, non-linear relationships better.
2. **Regularization in Ridge Regression**: **Ridge Regression** adds a regularization term to the linear regression equation, which helps control overfitting by penalizing large coefficients. This is important when you have many features that may not have significant predictive power. Regularization is especially useful when dealing with complex datasets where some features are correlated.
3. **Gradient Boosting’s Power**: **HistGradientBoosting** (or any gradient boosting model) is particularly good at capturing complex, non-linear relationships through boosting, which allows it to make improvements over simpler models. It also tends to handle missing values and outliers more effectively than traditional linear models.

Thus, **multilinear regression** could potentially work, but it may not offer the same flexibility and performance as the models you're currently using, especially with more complex datasets.