The evaluation of a "better" model depends on the specific goal and context of your analysis. Different metrics are used to assess different aspects of model performance, and the choice of which metric to prioritize depends on the problem you are trying to solve.

Here's a general guideline for common regression evaluation metrics:

**RMSE (Root Mean Square Error):** A lower RMSE indicates a better model. RMSE measures the average error between predicted and actual values, and lower values mean that the model's predictions are closer to the actual data points.

**MSE (Mean Square Error):** Similar to RMSE, a lower MSE is better. It measures the average squared difference between predicted and actual values. It is used when you want to emphasize the importance of larger errors in your evaluation.

**R-squared (R²) Score:** A higher R-squared value indicates a better model. R-squared measures the proportion of variance in the target variable explained by the model. An R-squared value closer to 1 indicates that the model explains a large portion of the variance, which is often desirable.

**Adjusted R-squared:** This is a modification of R-squared that adjusts for the number of predictors in the model. A higher adjusted R-squared indicates a better model, especially when comparing models with different numbers of features.

**Mean Absolute Error (MAE):** A lower MAE indicates a better model. MAE measures the average absolute difference between predicted and actual values.

**Mean Percentage Error (MPE)** or Mean Absolute Percentage Error (MAPE): These metrics are used when you want to measure the percentage difference between predicted and actual values. Lower values are better.

1. **statistic and p-value:** These metrics are used in the context of overall model significance. A lower p-value for the F-statistic suggests that the model as a whole is significant and better at explaining the variance in the data.

**Mean Absolute Error (MAE):**

The Mean Absolute Error (MAE) is another common metric used to measure the accuracy of a predictive model. It calculates the average of the absolute differences between predicted and actual values.

Ultimately, the choice of which metric to prioritize depends on your specific goals. For example, if your primary concern is prediction accuracy, you might focus on RMSE or MAE. If you're interested in understanding how much variance your model explains, R-squared or adjusted R-squared may be more relevant. Additionally, it's often a good practice to use a combination of these metrics and consider the context of your problem when determining what constitutes a "better" model.

**Proposed System Results.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MODEL** | **MAE** | **MSE** | **RMSE** | **R-squared (R²) Score** |
| Decision Tree Regressor | 0.5243564379960359 | 2.6402377695390324 | 1.6248808477974723 | 0.8132733063759845 |
| Random Forest | 0.528 | 3.008 | 1.734358671094304 | 0.7872638968727792 |
| Linear Regression | 0.6469544781787253 | 2.294108764277203 | 1.5146315605708218 | 0.8377527398064043 |
| Huber Regression | 0.36400001573554447 | 2.3239999991083664 | 1.5244671197203192 | 0.8356387288969487 |
| Least Angle Regression | 0.6469544781787256 | 2.294108764277203 | 1.5146315605708218 | 0.8377527398064043 |
| Lasso Regresion | 0.9345418157385402 | 2.639120238170266 | 1.5146315605708218 | 0.8133523420370595 |
| Lasso Least Angle Regression | 0.9345396967691519 | 2.63911656824673 | 1.62453580085104 | 0.8133526015866711 |
| **\***Dummy Regressor | 2.817227722772277 | 14.139588079600038 | 3.7602643629936496 | -2.8852334255091705e-07 |
| Ridge Regressor | 0.6470564484628811 | 2.294093600504422 | 1.5146265547997044 | 0.8377538122405566 |
| K-Nearest Neighbour | 0.6026666666666667 | 3.210765432098765 | 1.791860885252749 | 0.7729236282977798 |
| Elastic Net | 0.8984437391630721 | 2.563735265604195 | 1.6011668450240266 | 0.818683826511148 |
| Orthogonal Matching Pursuit | 0.6469544781787259 | 2.294108764277203 | 1.5146315605708218 | 0.8377527398064043 |
| LightGradient Boosting Matching | 0.652896612152291 | 3.927548055002587 | 1.9818042423515465 | 0.7222302965205634 |
| **MODEL** | **MAE** | **MSE** | **RMSE** | **R-squared (R²) Score** |
|  | 0.6473845598159994 | 2.2940470115940124 | 1.5146111750525322 | 0.8377571071684986 |
|  | 1.472015865776056 | 4.8015601973865625 | 2.1912462658009395 | 0.8293226838088108 |
| Catboost | 1.472015865776056 | 2.563212679425018 | 1.6010036475364502 | 0.8187207856026728 |
| Gradient boosting | 0.5771037785093528 | 2.5461309262309944 | 1.5956600283992184 | 0.8199288659248395 |
| Extra Tree Regresor | 1.5078975624751338 | 4.518142466225453 | 2.1255922624589725 | 0.6804614289765913 |
|  | 0.5959576568603515 | 2.5664476431310175 | 1.6020136213937188 | 0.8184919978458335 |

Mean Squared Error for xgb: 2.5664476431310175

Root Mean Squared Error for xgb: 1.6020136213937188

Mean Absolute Error for xgb: 0.5959576568603515

R-squared for xgb: 0.8184919978458335

Mean Absolute Percentage Error for xgb 0.020737961093790443

Explained Variance Score for xgb 0.8194462615085838

Median Absolute Error for xgb 0.14182281494140625

**Interpretation**

Let's interpret each of these metrics:

Mean Squared Error (MSE):

MSE measures the average squared difference between the predicted values and the actual values.

For the Dummy Regressor, the MSE of approximately 14.14 indicates that, on average, the squared difference between its predictions and the actual values is 14.14.

A higher MSE indicates a larger prediction error, which suggests that the Dummy Regressor's predictions are not accurate.

Root Mean Squared Error (RMSE):

RMSE is the square root of the MSE and is in the same units as the target variable.

In this case, the RMSE of approximately 3.76 means that, on average, the Dummy Regressor's predictions have an error of about 3.76 units.

A lower RMSE is generally preferred because it signifies smaller prediction errors.

Mean Absolute Error (MAE):

MAE measures the average absolute difference between the predicted values and the actual values.

The Dummy Regressor's MAE of approximately 2.82 indicates that, on average, its predictions have an absolute error of about 2.82 units.

R-squared (R²):

R-squared is a measure of how well the regression model explains the variance in the target variable. It ranges from -∞ to 1.

A value close to 1 indicates a good fit, while a value close to 0 suggests that the model does not explain much variance.

The Dummy Regressor's R² of approximately -2.89e-07 is very close to 0, indicating that it does not explain much variance and performs poorly.

Mean Absolute Percentage Error (MAPE):

MAPE measures the average percentage difference between the predicted values and the actual values.

The Dummy Regressor's MAPE of approximately 0.095 means that, on average, its predictions have an error of about 9.5% in relation to the actual values.

Explained Variance Score:

The explained variance score measures the proportion of the variance in the target variable that is explained by the model.

A score of 0.0 for the Dummy Regressor suggests that it does not explain any of the variance in the target variable.

Median Absolute Error:

Median Absolute Error measures the median absolute difference between the predicted values and the actual values.

The Dummy Regressor's median absolute error of approximately 0.618 indicates that, on average, its predictions have an absolute error of about 0.618 units.

In summary, the metrics collectively suggest that the Dummy Regressor is performing poorly and is not making accurate predictions. The MSE, RMSE, and MAE indicate relatively large prediction errors, while R-squared, explained variance, and MAPE show that the model does not explain much variance and has a high error percentage. These results emphasize the need for more advanced regression models to improve predictive accuracy for aspects of the menstrual cycle.

# Mean Squared Error (MSE)

mse = mean\_squared\_error(y\_test, y\_pred)

print("Mean Squared Error:", mse)

# Root Mean Squared Error (RMSE)

rmse = mean\_squared\_error(y\_test, y\_pred, squared=False)

print("Root Mean Squared Error:", rmse)

# Mean Absolute Error (MAE)

mae = mean\_absolute\_error(y\_test, y\_pred)

print("Mean Absolute Error:", mae)

# R-squared (R2)

r2 = r2\_score(y\_test, y\_pred)

print("R-squared:", r2)

# Mean Absolute Percentage Error (MAPE)

mape = mean\_absolute\_percentage\_error(y\_test, y\_pred)

print("Mean Absolute Percentage Error:", mape)

# Explained Variance Score

explained\_variance = explained\_variance\_score(y\_test, y\_pred)

print("Explained Variance Score:", explained\_variance)

# Median Absolute Error (MedAE)

medae = median\_absolute\_error(y\_test, y\_pred)

print("Median Absolute Error:", medae)

**Decision Tree Result**

Mean Squared Error: 2.6402377695390324

Root Mean Squared Error: 1.6248808477974723

Mean Absolute Error: 0.5243564379960359

R-squared: 0.8132733063759845

Mean Absolute Percentage Error: 0.017859589690949612

Explained Variance Score: 0.8135893622949752

Median Absolute Error: 0.0

Mean Squared Error (MSE): 2.6402

MSE measures the average squared difference between the predicted values and the actual (true) values.

In this case, an MSE of 2.6402 means that, on average, the squared difference between the model's predictions and the true values is approximately 2.64. Lower MSE values are better, indicating that the model's predictions are closer to the true values.

Root Mean Squared Error (RMSE): 1.6249

RMSE is the square root of the MSE and provides a measure of the average absolute error in the same units as the target variable.

In this case, an RMSE of 1.6249 means that, on average, the model's predictions are off by approximately 1.625 units. Again, lower RMSE values indicate better predictive performance.

Mean Absolute Error (MAE): 0.5244

MAE measures the average absolute difference between the predicted values and the actual values.

An MAE of 0.5244 means that, on average, the model's predictions are off by approximately 0.5244 units. MAE is also a measure of prediction accuracy, and lower values are preferred.

R-squared (R²): 0.8133

R-squared is a measure of how well the model explains the variance in the data. It ranges from 0 to 1, with higher values indicating a better fit.

In this case, an R² of 0.8133 means that the model explains approximately 81.33% of the variance in the data. It suggests that the model captures a significant portion of the variability in the target variable.

Mean Absolute Percentage Error (MAPE): 0.0179

MAPE measures the average percentage difference between the predicted values and the actual values.

An MAPE of 0.0179 means that, on average, the model's predictions deviate by approximately 1.79% from the true values as a percentage of the true values.

Explained Variance Score: 0.8136

The explained variance score is similar to R-squared and represents the proportion of the variance in the target variable that is explained by the model.

An explained variance score of 0.8136 suggests that approximately 81.36% of the variance in the target variable can be attributed to the model's predictions.

Median Absolute Error: 0.0

The median absolute error is the median of the absolute differences between predicted and true values.

A median absolute error of 0.0 indicates that, for at least half of the data points, the model's predictions are exactly equal to the true values, indicating a very accurate model for these data points.

In a nutshell, these metrics collectively provide insights into how well the model performs in terms of accuracy, precision, and the proportion of variance it can explain. In this case, the model appears to perform relatively well, as indicated by the low MSE, RMSE, MAE, MAPE, and the relatively high R-squared and explained variance score.

**Result for Random Forest**

窗体顶端

Mean Squared Error for Random Forest Algorithm: 3.008

Root Mean Squared Error for Random Forest Algorithm: 1.734358671094304

Mean Absolute Error for Random Forest Algorithm: 0.528

R-squared for Random Forest Algorithm: 0.7872638968727792

Mean Absolute Percentage Error for Random Forest Algorithm: 0.01719472155961226

Explained Variance Score for Random Forest Algorithm: 0.789259995202122

Median Absolute Error for Random Forest Algorithm: 0.0

**Result for Linear Regression**

Mean Squared Error for Linear Regression: 2.294108764277203

Root Mean Squared Error for Linear Regression: 1.5146315605708218

Mean Absolute Error for Linear Regression: 0.6469544781787253

R-squared for Linear Regression: 0.8377527398064043

Mean Absolute Percentage Error for Linear Regression: 0.02288643767416694

Explained Variance Score for Linear Regression: 0.8397249986717931

Median Absolute Error for Linear Regression: 0.36568428315927193

**Result for huber Regressor:**

Mean Squared Error for Huber Regressor: 2.3239999991083664

Root Mean Squared Error for Huber Regressor: 1.5244671197203192

Mean Absolute Error for Huber Regressor: 0.36400001573554447

R-squared for Huber Regressor: 0.8356387288969487

Mean Absolute Percentage Error for Huber Regressor: 0.0125682792420833

Explained Variance Score for Huber Regressor: 0.836590383206489

Median Absolute Error for Huber Regressor: 1.7545112029893062e-08

**Result for Least Angle Regressor**

Mean Squared Error for Least Angle Regressor: 2.294108764277203

Root Mean Squared Error for Least Angle Regressor: 1.5146315605708218

Mean Absolute Error for Least Angel Regressor: 0.6469544781787256

R-squared for Huber Regressor: 0.8377527398064043

Mean Absolute Percentage Error for Least Angel Regressor: 0.02288643767416696

Explained Variance Score for Least Angel Regressor: 0.8397249986717932

Median Absolute Error for Least Angel Regressor: 0.36568428315927193

**Result for Lasso Regression**

Mean Squared Error for Least Angle Regressor: 2.639120238170266

Root Mean Squared Error for Lasso Regression: 1.5146315605708218

Mean Absolute Error for Lasso Regression: 0.9345418157385402

R-squared for Lasso Regression: 0.8133523420370595

Mean Absolute Percentage Error for Lasso Regression: 0.032339144072889736

Explained Variance Score for Lasso Regresion: 0.8145772491957535

Median Absolute Error for Lasso Regression: 0.6179639164098845

**Result for Lasso Least Angel Regression**

Mean Squared Error for Lasso Least Angle Regressor: 2.63911656824673

Root Mean Squared Error for Lasso Least Angle Regression: 1.62453580085104

Mean Absolute Error for Lasso Least Angel Regression: 0.9345396967691519

R-squared for Lasso Least Angel Regression: 0.8133526015866711

Mean Absolute Percentage Error for Lasso Least Angel Regression: 0.03233907309692495

Explained Variance Score for Lasso Least Regresion: 0.8145775106114873

Median Absolute Error for Lasso Least Angle Regression: 0.6179607175297441

**Results of Extracting the selected features and their coefficients.**

Selected Features (LassoLars):

EstimatedDayofOvulation

LengthofLutealPhase

Coefficients (LassoLars):

EstimatedDayofOvulation: 0.8066329

LengthofLutealPhase: 0.72212832

Interpreting these results:

**Result for Light Gradient Boosting**

Mean Squared Error for LightGBM: 3.927548055002587

Root Mean Squared Error for LightGBM: 1.9818042423515465

Mean Absolute Error for y\_pred\_LightGBM: 0.652896612152291

R-squared for LightGBM: 0.7222302965205634

Mean Absolute Percentage Error for LightGBM 0.021990485436275783

Explained Variance Score for LightGBM: 0.7239652896162208

Median Absolute Error for LightGBM 0.012535964766136942

Estimated Day of Ovulation (EstimatedDayofOvulation): This feature has a positive coefficient of approximately 0.807. In the context of modeling the menstrual cycle, this suggests that as the estimated day of ovulation increases, it has a positive impact on the dependent variable (possibly cycle length or some other related factor). In other words, later estimated days of ovulation are associated with longer menstrual cycles or other changes in the menstrual cycle.

**Result for Dummy Regressor**

Mean Squared Error for Dummy Regressor: 14.139588079600038

Root Mean Squared Error for Dummy Regressor: 3.7602643629936496

Mean Absolute Error for Dummy Regressor: 2.817227722772277

R-squared for Dummy Regressor: -2.8852334255091705e-07

Mean Absolute Percentage Error for Dummy Regressor 0.0947626266655161

Explained Variance Score for Dummy Regressor: 0.0

Median Absolute Error for Dummy Regressor: 2.301980198019802

Length of Luteal Phase (LengthofLutealPhase): This feature also has a positive coefficient of approximately 0.722. The luteal phase is the second half of the menstrual cycle, which occurs after ovulation. A longer luteal phase is associated with a later estimated day of ovulation, and this feature also contributes positively to the dependent variable, suggesting that a longer luteal phase is associated with changes in the menstrual cycle.

**Result for ridge Regressor**

Mean Squared Error for ridge Regressor: 2.294093600504422

Root Mean Squared Error for Ridge Regressor: 1.5146265547997044

Mean Absolute Error for Ridge Regressor: 0.6470564484628811

R-squared for Ridge Regressor: 0.8377538122405566

Mean Absolute Percentage Error for Ridge Regressor 0.02288993783118628

Explained Variance Score for Ridge Regressor: 0.8397255211274202

Median Absolute Error for Ridge Regressor: 0.36577863991946913

**Result for KNN**

Mean Squared Error for KNN: 3.988640000000001

Root Mean Squared Error for KNN: 1.9971579807316198

Mean Absolute Error for KNN: 0.6568000000000002

R-squared for KNN: 0.717909664103272

Mean Absolute Percentage Error for KNN 0.021361321891354302

Explained Variance Score for KNN: 0.7299145434547437

Median Absolute Error for KNN: 0.0

**Result for Elastic Net**

Mean Squared Error for Elastic Net: 2.563735265604195

Root Mean Squared Error for Elastic Net: 1.6011668450240266

Mean Absolute Error for Elastic Net: 0.8984437391630721

R-squared for Elastic Net: 0.818683826511148

Mean Absolute Percentage Error for Elastic Net 0.03115830932611471

Explained Variance Score for Elastic Net: 0.8199882890156751

Median Absolute Error for Elastic Net 0.5653998842733898

**Result for Orthogonal Matching Pursuit**

Mean Squared Error for Orthogonal Matching Pursuit: 2.294108764277203

Root Mean Squared Error for Orthogonal Matching Pursuit: 1.5146315605708218

Mean Absolute Error for Othogonal Matching Pursuit: 0.6469544781787259

R-squared for Othogonal Matching Pursuit: 0.8377527398064043

Mean Absolute Percentage Error for Othogonal Matching Pursuit 0.02288643767416697

Explained Variance Score for orthogonal Matchiing Pursuit: 0.8397249986717932

Median Absolute Error for Orthogonal Matching Pursuit 0.36568428315927193

**Result for Bayelsian Ridge Regressor**

Mean Squared Error for Bayesian Ridge: 2.2940470115940124

Root Mean Squared Error for Bayesian Ridge: 1.5146111750525322

Mean Absolute Error for Bayesian Ridge: 0.6473845598159994

R-squared for Bayesian Ridge: 0.8377571071684986

Mean Absolute Percentage Error for Bayesian Ridge 0.021990485436275783

Explained Variance Score for Bayesian Ridge: 0.8397270471039116

Median Absolute Error for Bayesian Ridge 0.3660822450197436

**RESULT FOR PASSIVE AGGRESSIVE**

Mean Squared Error for passive Aggressive: 2.4133062491798807

Root Mean Squared Error for Passive Aggressive: 1.5534819758142933

Mean Absolute Error for Passive Aggressive: 0.5677120652915696

R-squared for Passive Aggressive: 0.8293226838088108

Mean Absolute Percentage Error for Passive Aggressive 0.01913918649368607

Explained Variance Score for Pred\_passive Aggressive: 0.8330599724141792

Median Absolute Error for Passive Aggressive 0.17607529082654594

**Result for Adaboost**

Mean Squared Error for Adaboost Regressor : 4.8015601973865625

Root Mean Squared Error for Adaboost: 2.1912462658009395

Mean Absolute Error for Passive Aggressive: 1.472015865776056

R-squared for Adaboost : 0.8293226838088108

Mean Absolute Percentage Error for Adaboost 0.0516084488582049

Explained Variance Score for Adaboost 0.6713029512339932

Median Absolute Error for Adaboost 1.0538418591667256

**Result for Gradient Boosting**

Mean Squared Error for Gradient\_boosting: 2.5461309262309944

Root Mean Squared Error for Gradient\_boosting: 1.5956600283992184

Mean Absolute Error for Gradient\_boosting: 0.5771037785093528

R-squared for Gradient\_boosting: 0.8199288659248395

Mean Absolute Percentage Error for Gradient\_boosting 0.020000598291770002

Explained Variance Score for Gradient\_boosting 0.8206343253216919

Median Absolute Error for Gradient\_boosting 0.1250994890651338

**Result for Catboost**

Mean Squared Error for Catboost : 2.563212679425018

Root Mean Squared Error for Catboost: 1.6010036475364502

Mean Absolute Error for Catboost: 1.472015865776056

R-squared for Catboost: 0.8187207856026728

Mean Absolute Percentage Error for Catboost 0.016378396450027893

Explained Variance Score for Catboost 0.8193694989790116

Median Absolute Error for Catboost 0.012537043901616585

**Result for Extra Tree Regressor**

Mean Squared Error for Extra trees: 4.518142466225453

Root Mean Squared Error for Extra Trees: 2.1255922624589725

Mean Absolute Error for Extra Trees: 1.5078975624751338

R-squared for Extra Trees: 0.6804614289765913

Mean Absolute Percentage Error for Extra Trees 0.05203805856918657

Explained Variance Score for Extra Trees 0.6811362548885502

Median Absolute Error for Extra Trees 1.0591676641879104

**Result for Xgb**

Mean Squared Error for xgb: 2.5664476431310175

Root Mean Squared Error for xgb: 1.6020136213937188

Mean Absolute Error for xgb: 0.5959576568603515

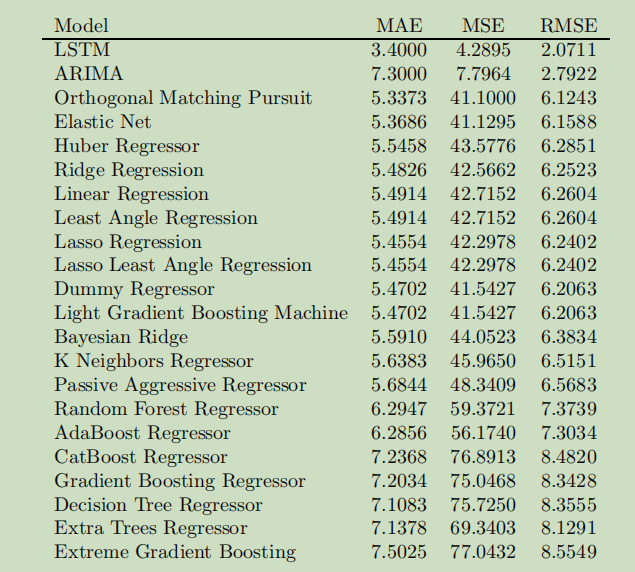
R-squared for xgb: 0.8184919978458335

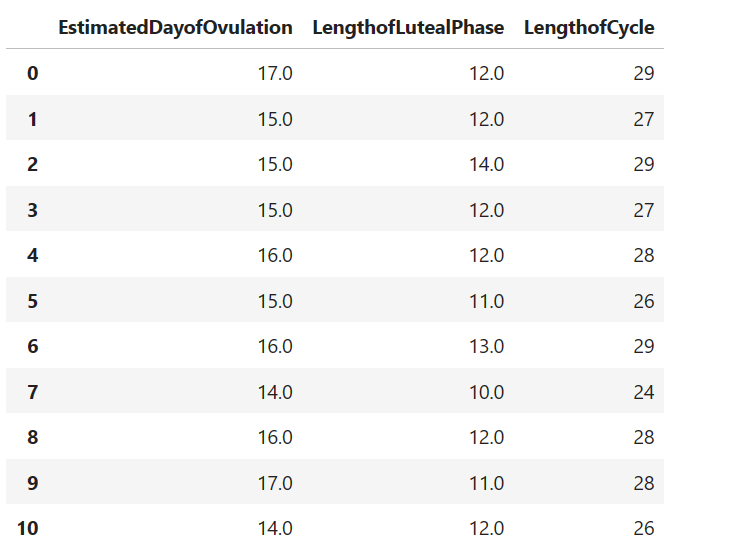
Mean Absolute Percentage Error for xgb 0.020737961093790443

Explained Variance Score for xgb 0.8194462615085838

Median Absolute Error for xgb 0.14182281494140625

**Existing System**





Based on the above menstrual cycle information, the decision tree is giving accurate prediction

