Discussion @Melody

In this section, it will interpret the models and discuss the significance of the results. **See the Regression Models table below.**

In the Linear Regression Model, the correlation coefficients of each feature are low which suggests that some of them may have a weak relation with kda. To be specific, it could be getting the Train (R2) and Test (R2) are both low since the best possible score should be 1.0. Besides, The RMSE value is not large but still shows that there could be ±2.6244 value between the predict and true value. Therefore, the Linear Regression cannot be the satisfied model.

For decision tree regression, it could be found out that Test (R2) is negative after doing 5-fold CV which represents that the model is worse to predict than using the average. Moreover, the Train (R2) value is 1.0 which also has large difference with 5-fold CV (R2). However, through MSE, the regression tree performs even worse than the linear regression models. Therefore, decision tree regression model also cannot have statistical significance.

Consider the last model random forest regression, the Train (R2) value is 0.8888 which improves a lot as well as becomes closer to the expected predictor. In addition, compared with the previous 2 models, the RMSE value in random forest regression is lower which also represents that there has smaller value difference between the predicted value and the true value.

In conclusion, it is not hard to find out that the random forest regression model is the most meaningful model that could better demonstrate the strong connections between variables and kda.

*Table: Regression Models*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Train (R2) | Test (R2) | 5-fold CV (R2) | MSE | RMSE |
| Linear Regression | 0.1689 | 0.1934 | 0.1643 | 6.8879 | 2.6244 |
| Decision Tree Regression | 1.0 | -0.4231 | -0.5509 | 12.1534 | 3.4861 |
| Random Forest Regression | 0.8888 | 0.2526 | 0.2082 | 6.3861 | 2.5271 |

Below the figure shows that the relationships between feature importance and the mean decrease in impurity. Simply can get that the top three high relative factors except kills, death and assists are damage\_building, damage\_total and gold\_earned, while damage\_building has the highest value 0.1725 which represents that it has the strongest connection to kda than any other features. Meanwhile, the feature level has the lowest value 0.044 which performs weak relationships with kda.

图表, 条形图

描述已自动生成

*Figure :* Feature importance with mean decrease in impurity.