Index: 210170G
Name: W.A.R.T. Fonseka

Lab 02 – Regression

Report

Introduction:

In this report, I aim to find the best prediction data for habitability scores by exploring various machine learning approaches. Furthermore, this report discusses various evaluation metrics for ML regression scenarios and attempts to find the most suitable evaluation metric. We also discuss various optimization techniques for ML models.

Evaluation matrices:

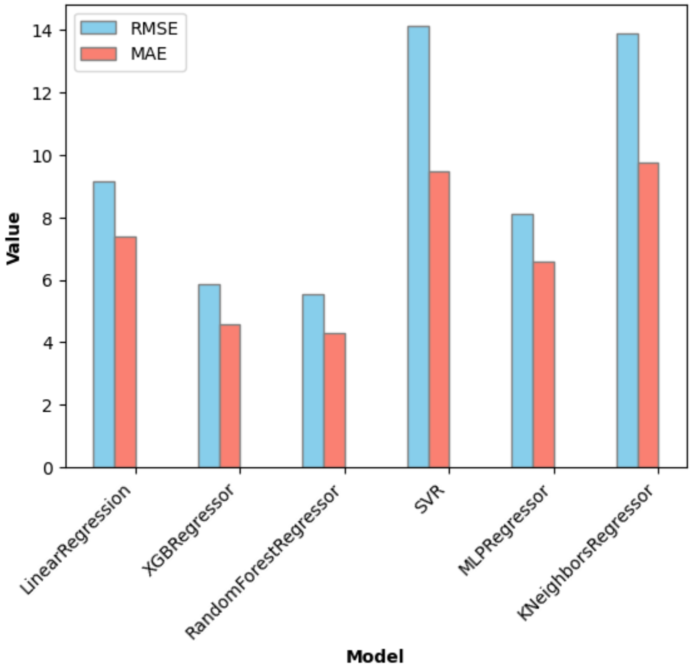
I used the following machine learning models to evaluate and select the best-performing model in this scenario:

1. XGB Regressor
2. Random Forest Regressor
3. SVR
4. Linear Regression
5. MLP Regressor
6. K Neighbors Regressor

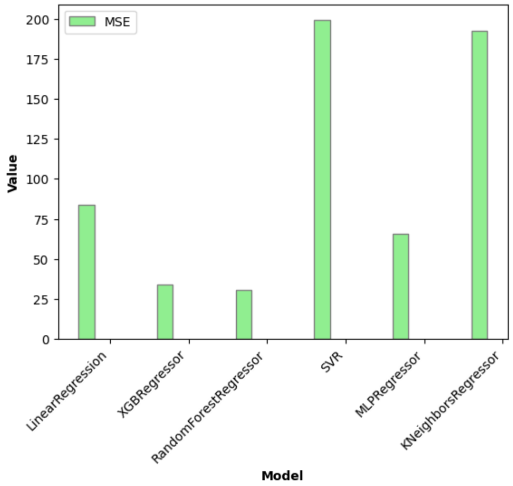
And I used the following evaluation metrics to assess the performance of each model:

1. Root mean squared error
2. Mean absolute error
3. Mean squared error
4. r2 score

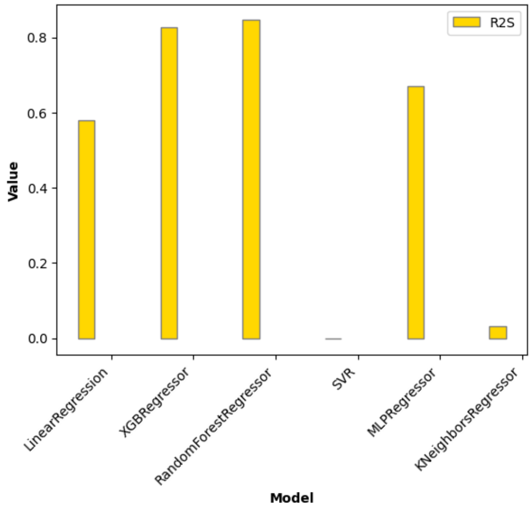
Below is the chart showing the values of the evaluation metrics:



We can clearly see that XGB Regressor and Random Forest Regressor give the minimum Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The third best model is the Linear Regression model. SVR, MLP Regressor, and K Neighbors Regressor show significantly higher error values compared to the other three models. Additionally, we observe a similar trend in RMSE and MAE. As the RMSE increases or decreases for a particular model, the MAE also tends to follow a similar pattern. This indicates a consistent relationship between the two evaluation metrics across the models being compared.



Similarly, we can observe that Mean Square Error (MSE) follows a similar trend. However, the values are too high compared to RMSE and MAE due to its squaring operation.



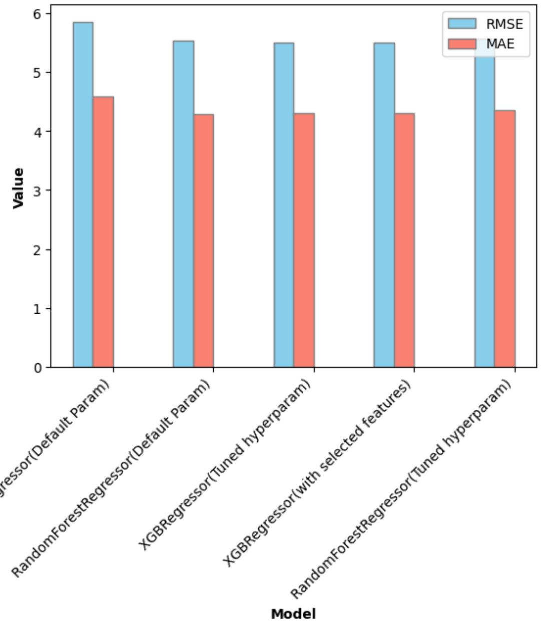
In R2S, we can observe an inverse proportion trend compared to the above three metrics. However, it signifies the same trend, as R2S indicates how well the model explains variability in the data. A high R2S value means that more of the data can be explained by the model, while a low R2S value indicates less data can be explained by the model, other words poor performance.

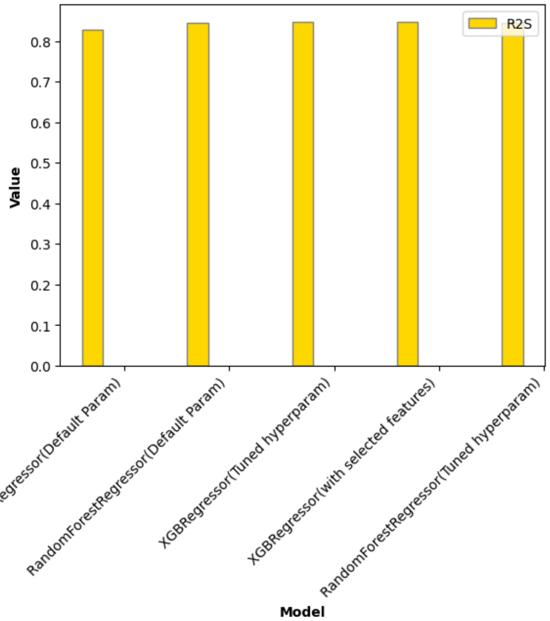
All four metrics indicate that XGB Regressor and Random Forest Regressor perform better on this data compared to the others.

Different approaches:

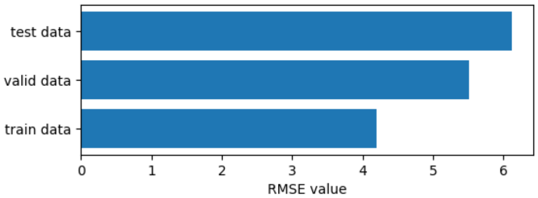
I use three different approaches, they are:

1. Single model(XGB Regressor and Random Forest Regressor with default parameters, tuned parameters and with best selected features)





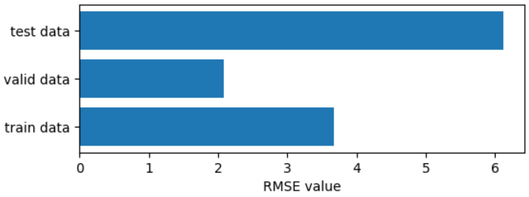
We can observe that the RMSE, MAE, and R2S values of all single models are more similar.



This chart indicates the performance of the XGB Regressor model with selected features on the training, validation, and test datasets.

RMSE value for test data is 6.34

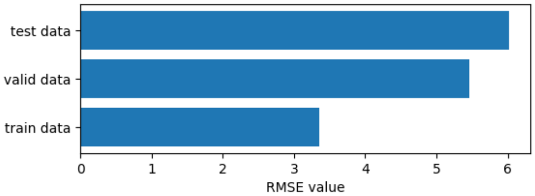
2. Ensemble method with Blending



This chart indicates the performance of the Ensemble model with blending on the training, validation, and test datasets. Model well perform on valid data because of ensemble model train on it. But not reduce error on other test data set.

RMSE value for test data is 6.22

3. Ensemble method with Averaging

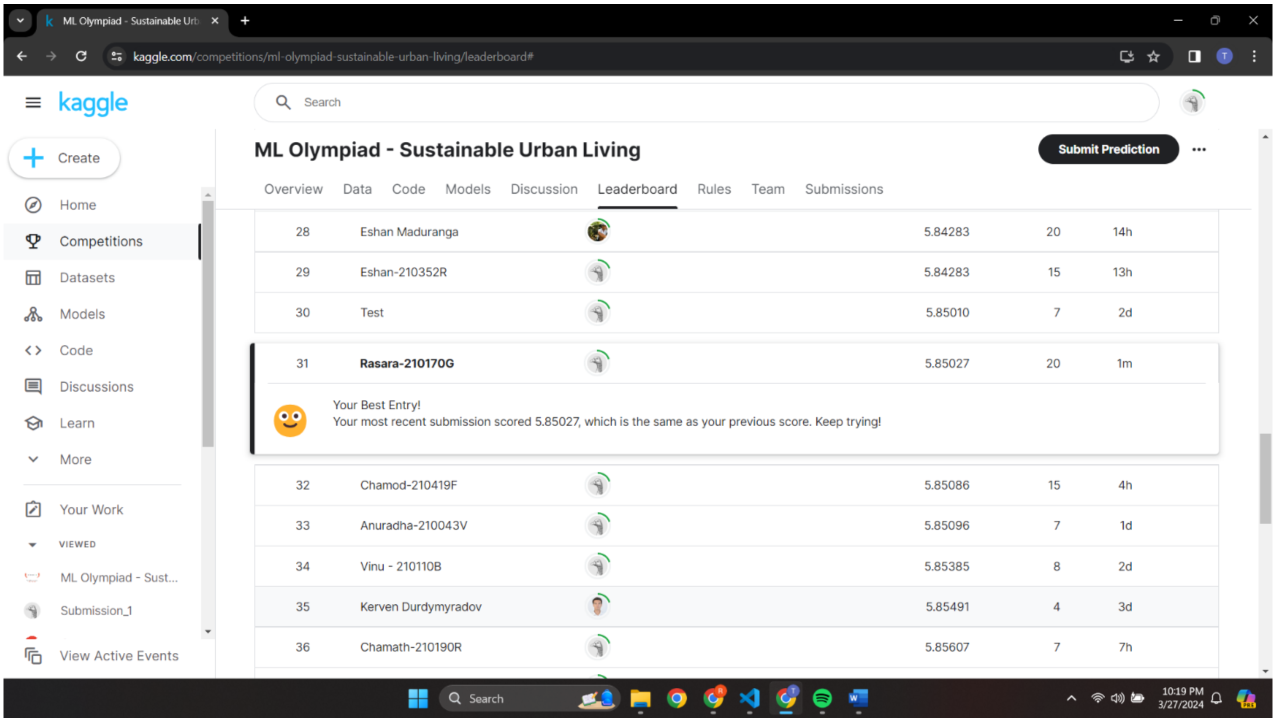


This chart indicates the performance of the Ensemble model with Averaging on the training, validation, and test datasets. Also it's doesn't show significant improvement.

RMSE value for test data is 6.02

Since Ensemble method with Averaging model's error lower than the Single XGB Regressor model's error by 5.32% and it's also lower than the Blending model's error by 3.32%.
So, I used Averaging Ensemble model as my final model and trained it all the given data(train + valid) and predict Habitability score of test data set.

My final RMSE for test data is 5.85.
It's became low because more data trains on it.



ML Olympiad - Sustainable Urban Living

28	Eshan Maduranga	5.84283	20	14h
29	Eshan-210352R	5.84283	15	13h
30	Test	5.85010	7	2d
31	**Rasara-210170G**	5.85027	20	1m

Your Best Entry!
Your most recent submission scored 5.85027, which is the same as your previous score. Keep trying!

32	Chamod-210419F	5.85086	15	4h
33	Anuradha-210043V	5.85096	7	1d
34	Vinu - 210110B	5.85385	8	2d
35	Kerven Durdymyradov	5.85491	4	3d
36	Chamath-210190R	5.85607	7	7h