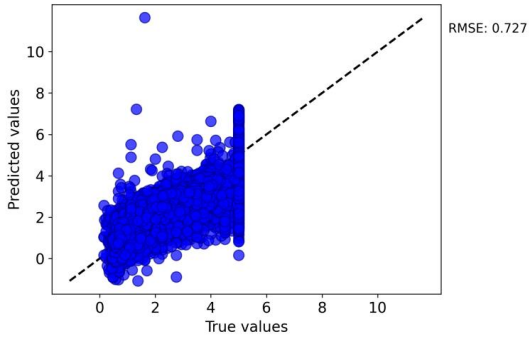
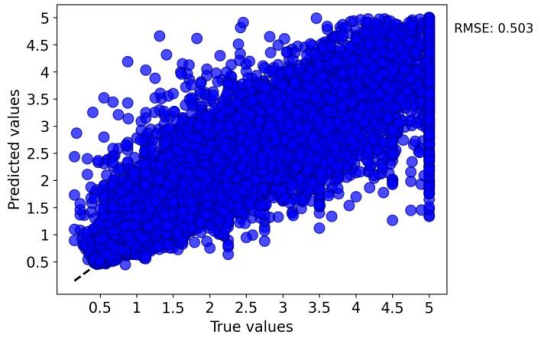
Assessment Figures

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude
0	2.344766	0.982143	0.628559	-0.153758	-0.974429	-0.049597	1.052548	-1.327835
1	2.332238	-0.607019	0.327041	-0.263336	0.861439	-0.092512	1.043185	-1.322844
2	1.782699	1.856182	1.155620	-0.049016	-0.820777	-0.025843	1.038503	-1.332827
3	0.932968	1.856182	0.156966	-0.049833	-0.766028	-0.050329	1.038503	-1.337818
4	-0.012881	1.856182	0.344711	-0.032906	-0.759847	-0.085616	1.038503	-1.337818
...	...	...	...	...	...	...	...	...
20635	-1.216128	-0.289187	-0.155023	0.077354	-0.512592	-0.049110	1.801647	-0.758826
20636	-0.691593	-0.845393	0.276881	0.462365	-0.944405	0.005021	1.806329	-0.818722
20637	-1.142593	-0.924851	-0.090318	0.049414	-0.369537	-0.071735	1.778237	-0.823713
20638	-1.054583	-0.845393	-0.040211	0.158778	-0.604429	-0.091225	1.778237	-0.873626
20639	-0.780129	-1.004309	-0.070443	0.138403	-0.033977	-0.043682	1.750146	-0.833696

20640 rows × 8 columns





ML4ER - Assignment 4 Activities

Muhammad Zain Azeem,
Informatics Skunkworks (**non-credits),** Week 2
01/08/2024

Progress

Dataset Normalization

StandardScaler()

- mean & standard deviation

Examples:

```
[27] X.AveRooms.mean()
     6.609699867535816e-17

[28] X.AveRooms.std()
     1.0000242256864988
```

Normalization Results

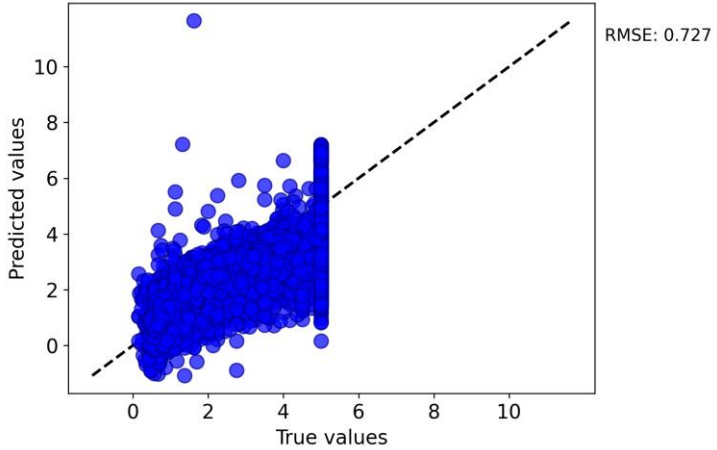
	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude
0	2.344766	0.982143	0.628559	-0.153758	-0.974429	-0.049597	1.052548	-1.327835
1	2.332238	-0.607019	0.327041	-0.263336	0.861439	-0.092512	1.043185	-1.322844
2	1.782699	1.856182	1.155620	-0.049016	-0.820777	-0.025843	1.038503	-1.332827
3	0.932968	1.856182	0.156966	-0.049833	-0.766028	-0.050329	1.038503	-1.337818
4	-0.012881	1.856182	0.344711	-0.032906	-0.759847	-0.085616	1.038503	-1.337818
...	...	...	...	...	...	...	...	...
20635	-1.216128	-0.289187	-0.155023	0.077354	-0.512592	-0.049110	1.801647	-0.758826
20636	-0.691593	-0.845393	0.276881	0.462365	-0.944405	0.005021	1.806329	-0.818722
20637	-1.142593	-0.924851	-0.090318	0.049414	-0.369537	-0.071735	1.778237	-0.823713
20638	-1.054583	-0.845393	-0.040211	0.158778	-0.604429	-0.091225	1.778237	-0.873626
20639	-0.780129	-1.004309	-0.070443	0.138403	-0.033977	-0.043682	1.750146	-0.833696

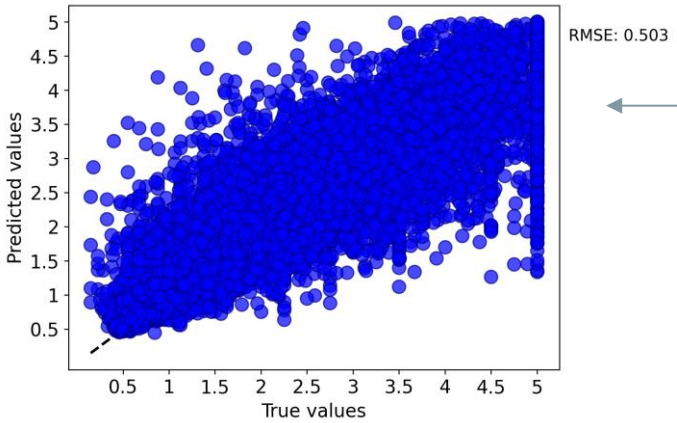
20640 rows × 8 columns

Progress

Model Fitting

- Linear Regression Model
- Kernel Ridge Model





Think about the following:

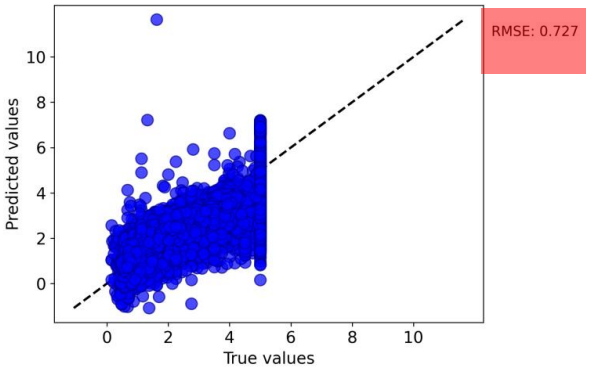
Compare the qualitative performance of each model via your parity plots. Which model is performing better?

do the same with the reported error metrics. Which model is performing better?

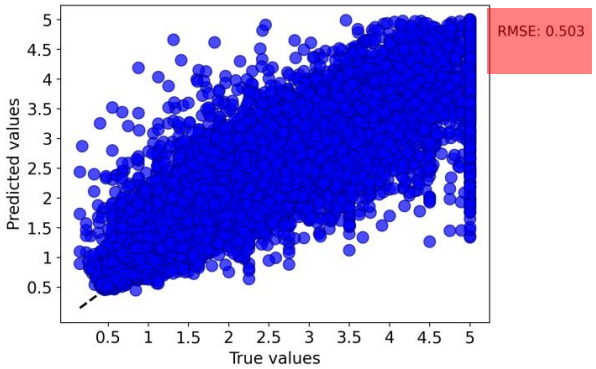
- Kernel Ridge Model due to **lower RMSE value**!

Progress

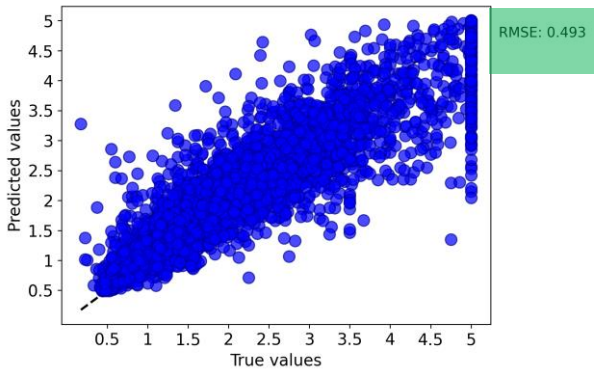
Linear Regression Model



Kernel Ridge Model



RandomForestRegressor



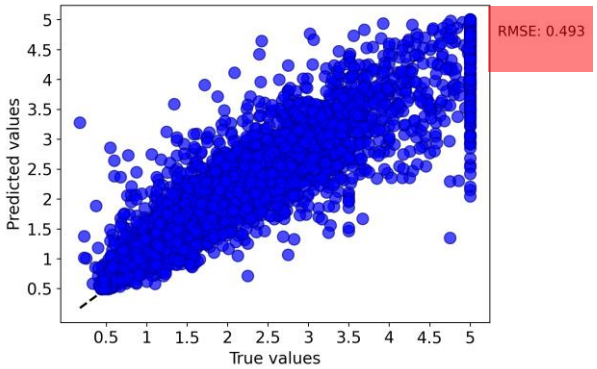
Conclusion

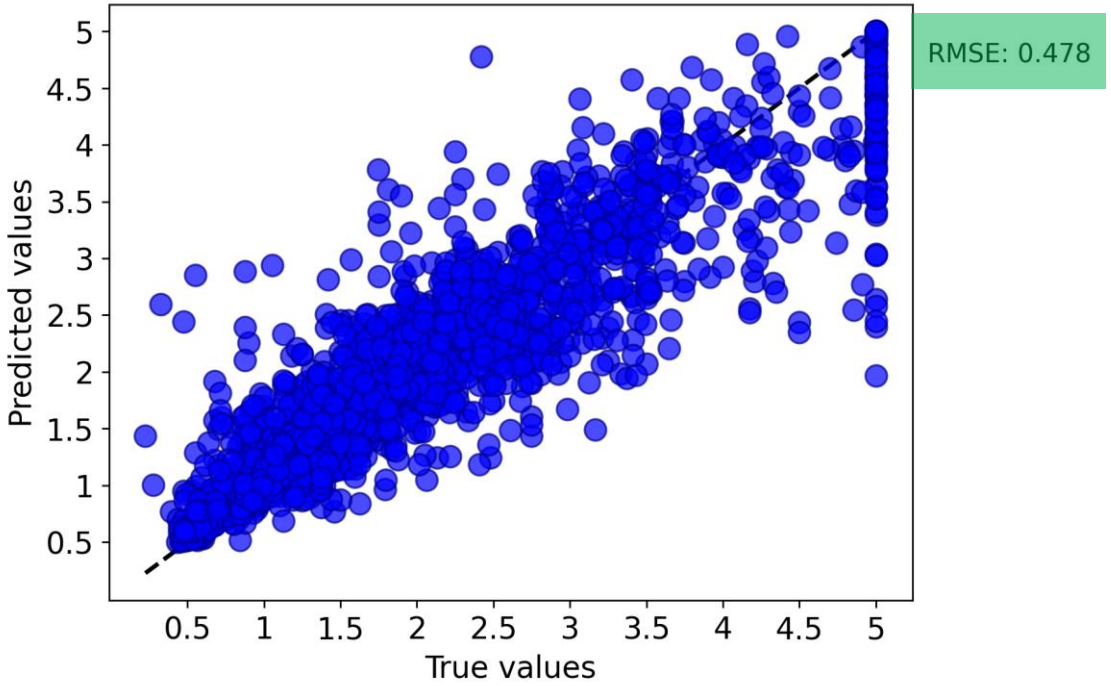
RandomForestRegressor was found to be **effective**!

* 5-fold CV for each model

Progress

RandomForestRegressor (5-fold CV)





RandomForestRegressor (10-fold CV)