Boston Housing Prediction Results

Row	Prediction Model	Feature Selection Method	Feature removed	Normalization Method	Train Test Method	Test Size	Train Score	Test Score	Train & Test Score (Weighted AVG)	Model Rank
1	Linear Regression	-	-	-	70% Train_30% Test	30%	87%	-722%	-156%	15
2	Linear Regression	-	-	Standard Scalar	70% Train_30% Test	30%	87%	-722%	-156%	15
3	Linear Regression	-	-	-	Train_Test_Split	30%	76%	67%	73%	7
4	Linear Regression	-	-	Standard Scalar	Train_Test_Split	30%	76%	67%	73%	7
5	Linear Regression	Correlations (>0.7 or <-0.7)	'INDUS', 'DIS', 'NOX', 'TAX'	-	70% Train_30% Test	30%	85%	-257%	-18%	13
6	Linear Regression	Correlations (>0.7 or <-0.7)	'INDUS', 'DIS', 'NOX', 'TAX'	Standard Scalar	70% Train_30% Test	30%	85%	-257%	-18%	13
7	Linear Regression	Correlations (>0.7 or <-0.7)	'INDUS', 'DIS', 'NOX', 'TAX'	-	Train_Test_Split	30%	73%	62%	70%	10
8	Linear Regression	Correlations (>0.7 or <-0.7)	'INDUS', 'DIS', 'NOX', 'TAX'	Standard Scalar	Train_Test_Split	30%	73%	62%	70%	10
9	Linear Regression	Chi2	'NOX', 'RM', 'PTRATIO', 'CHAS'	-	70% Train_30% Test	30%	60%	-1118%	-293%	17
10	Linear Regression	Chi2	'NOX', 'RM', 'PTRATIO', 'CHAS'	-	Train_Test_Split	30%	67%	58%	64%	12
11	Random Forest Regressor	Chi2	'NOX', 'RM', 'PTRATIO', 'CHAS'	Standard Scalar	Train_Test_Split	30%	98%	72%	90%	3
12	Random Forest Regressor	Chi2	'NOX', 'RM', 'PTRATIO', 'CHAS'	-	Train_Test_Split	30%	98%	71%	90%	4
13	Random Forest Regressor	Correlations (>0.7 or <-0.7)	'INDUS', 'DIS', 'NOX', 'TAX'	-	Train_Test_Split	30%	98%	76%	91%	2
14	Random Forest Regressor	-	-	Standard Scalar	Train_Test_Split	30%	98%	84%	94%	1
15	Random Forest Regressor	-	-	Standard Scalar	70% Train_30% Test	30%	99%	24%	77%	6
16	Plynomial (Degree 2)	Chi2	'NOX', 'RM', 'PTRATIO', 'CHAS'	Standard Scalar	Train_Test_Split	30%	80%	53%	72%	9
17	Plynomial (Degree 2)	Correlations (>0.7 or <-0.7)	'INDUS', 'DIS', 'NOX', 'TAX'	Standard Scalar	Train_Test_Split	30%	89%	51%	78%	5
18	Plynomial (Degree 3)	Correlations (>0.7 or <-0.7)	'INDUS', 'DIS', 'NOX', 'TAX'	Standard Scalar	Train_Test_Split	30%	96%	-7822%	-2279%	18
19										
20										

Description :

1- One of the most important reasons for the difference between the scores of the models is the **training and testing method** **(Rows : 1 , 3 or 14 , 15)**.

2- **Normalization** of the available samples makes very little difference in the final results **(Rows : 1 , 2)**.

3 - In a specified model, assuming the stability of other conditions, the **removal of features** in various methods, contrary to expectations, **reduces the performance** of the prediction model! **(Rows : 3 , 7 , 10)**.

4 - **Increasing the degree in the polynomial method** increases the consistency of the training part, but at the same time greatly reduces the test result. In a way, it can be said that Increasing the degree will result model suffers from **overfit** **(Rows : 17 , 18)**.