AI Project :
House Price Prediction

112252 임우담
15010974 이혜민
16011106 조경빈

INDEX

House Price Prediction

1 Data Set

1) Types of Features

- Train Data Size : (1460, 80)
- Test Data Size : (1459, 79)

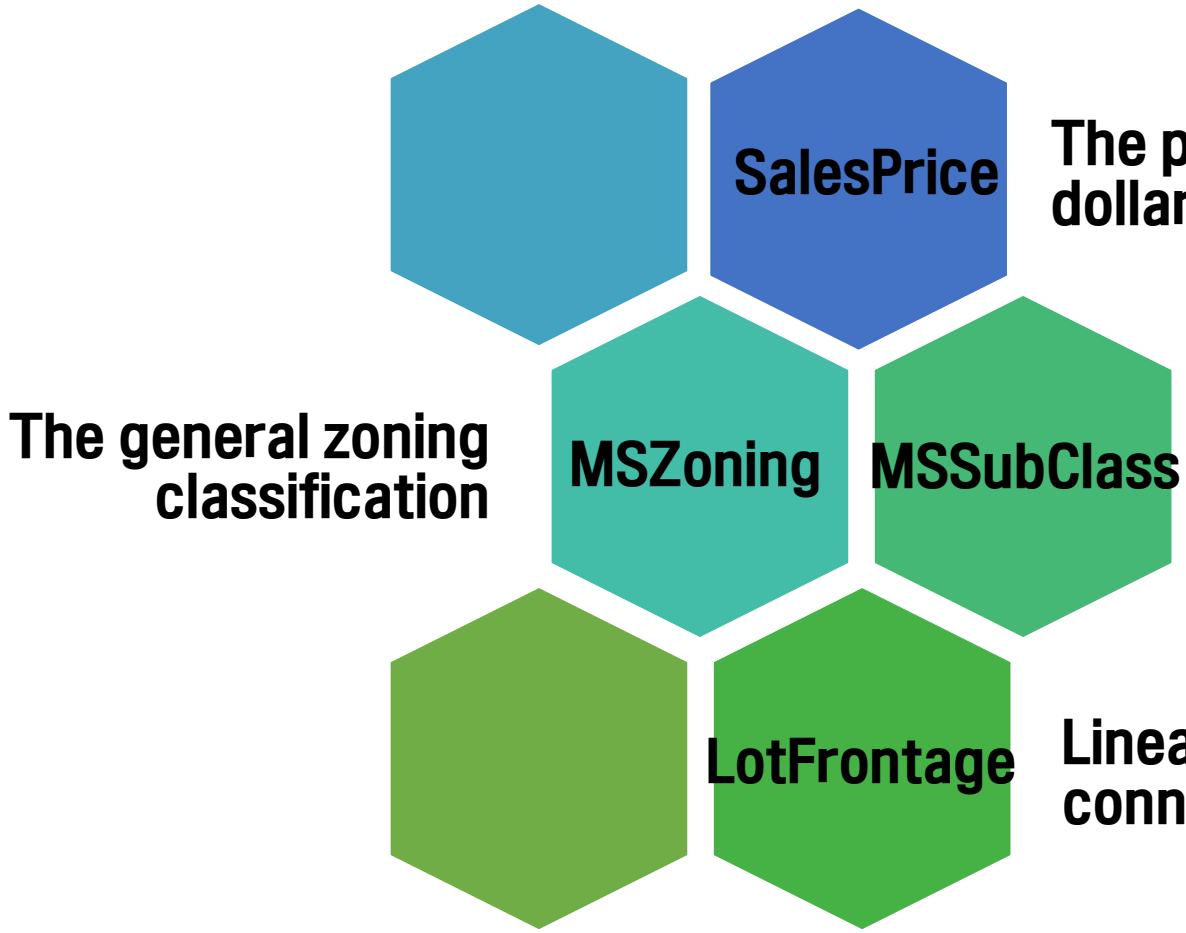
	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...
0	1461	20	RH	80.000	11622	Pave	NaN	Reg	Lvl	AllPub	...
1	1462	20	RL	81.000	14267	Pave	NaN	IR1	Lvl	AllPub	...
2	1463	60	RL	74.000	13830	Pave	NaN	IR1	Lvl	AllPub	...
3	1464	60	RL	78.000	9978	Pave	NaN	IR1	Lvl	AllPub	...
4	1465	120	RL	43.000	5005	Pave	NaN	IR1	HLS	AllPub	...

5 rows × 80 columns

ScreenPorch	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondition
120	0	NaN	MnPrv	NaN	0	6	2010	WD	Normal
0	0	NaN	NaN	Gar2	12500	6	2010	WD	Normal
0	0	NaN	MnPrv	NaN	0	3	2010	WD	Normal
0	0	NaN	NaN	NaN	0	6	2010	WD	Normal
144	0	NaN	NaN	NaN	0	1	2010	WD	Normal

1 Data Set

1) Types of Features



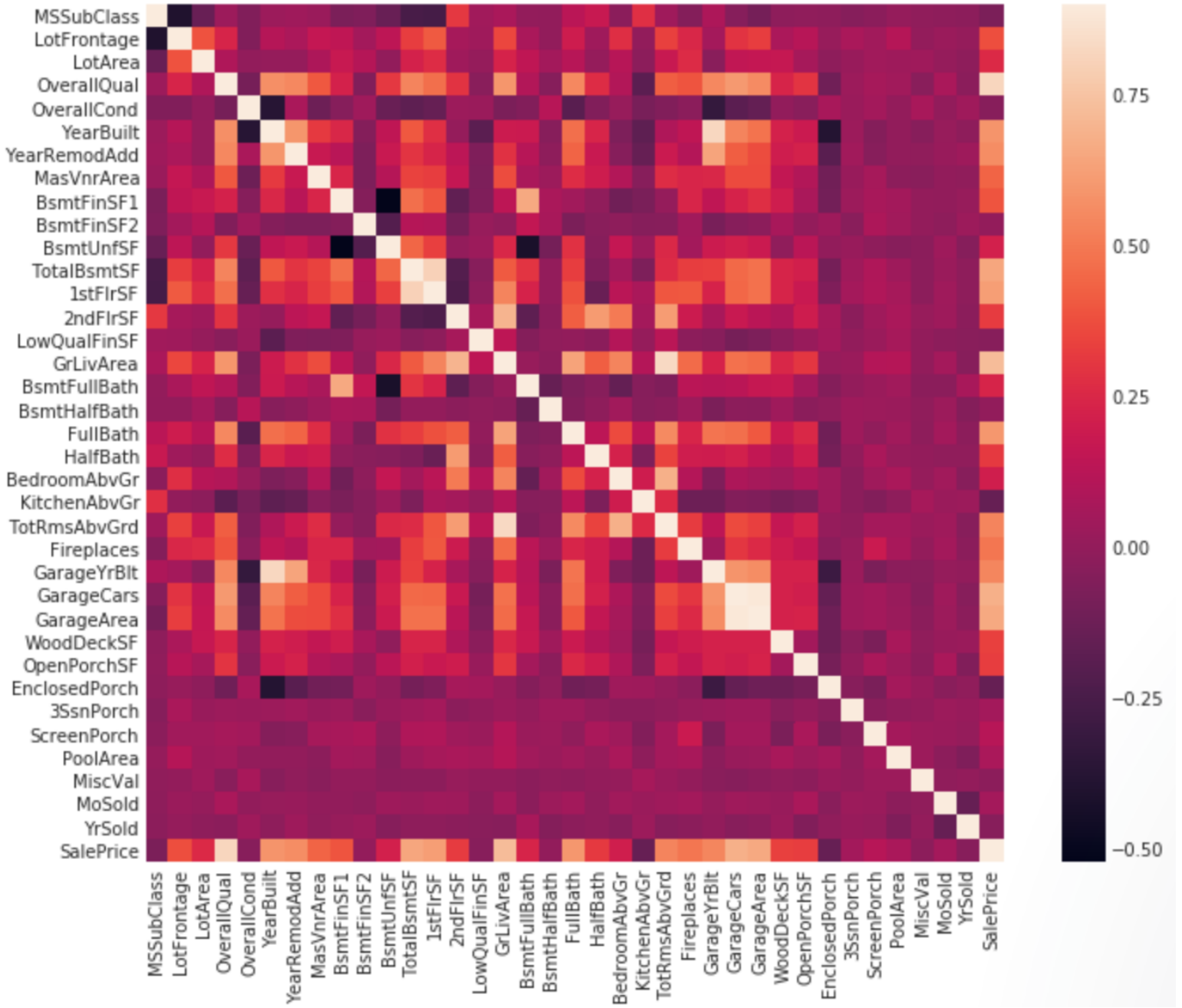
1 Data Set

2) Data Correlation

→ Correlation map to see how features are correlated with SalePrice

Heatmap

1


MSSubClass
LotFrontage
LotArea
OverallQual
OverallCond
YearBuilt
YearRemodAdd
MasVnrArea
BsmtFinSF1
BsmtFinSF2
BsmtUnfSF
TotalBsmtSF
1stFlrSF
2ndFlrSF
LowQualFinSF
GrLivArea
BsmtFullBath
BsmtHalfBath
FullBath
HalfBath
BedroomAbvGr
KitchenAbvGr
TotRmsAbvGrd
Fireplaces
GarageYrBlt
GarageCars
GarageArea
WoodDeckSF
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea
MiscVal
MoSold
YrSold
SalePrice
0.75
0.50
0.25
0.00
−0.25
−0.50

1 Data Set

2) Data Correlation

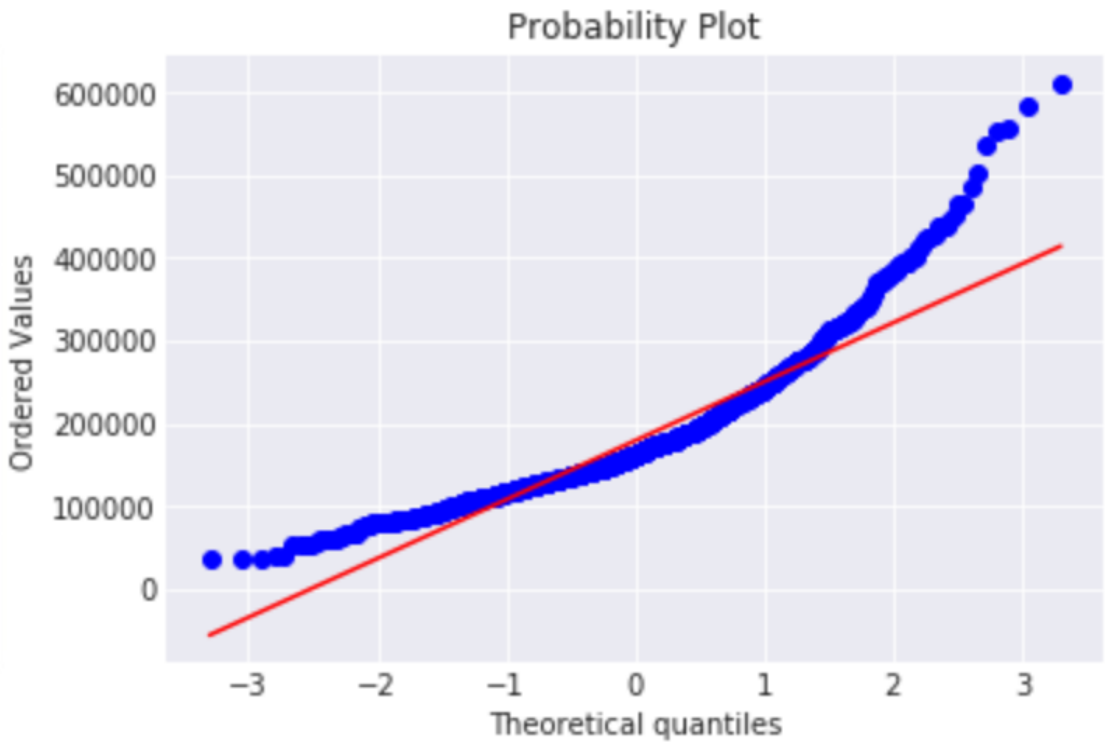
→ Top 9

OverallQual, GrLivArea, GarageCars, 1stFlrSF, YearBuilt, FullBath, TotRmsAbvGrd, GarageArea, Fireplaces

1 Data Set

3) Data Analysis

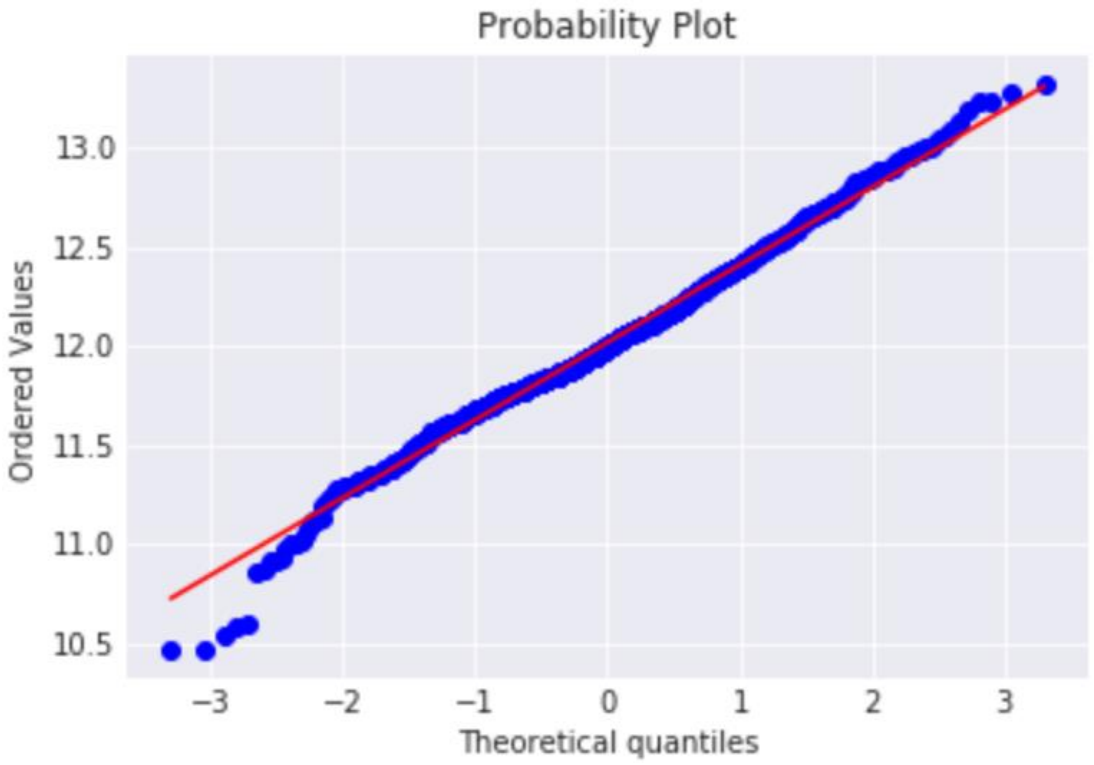
- SalesPrice : Before Normalization



1 Data Set

3) Data Analysis

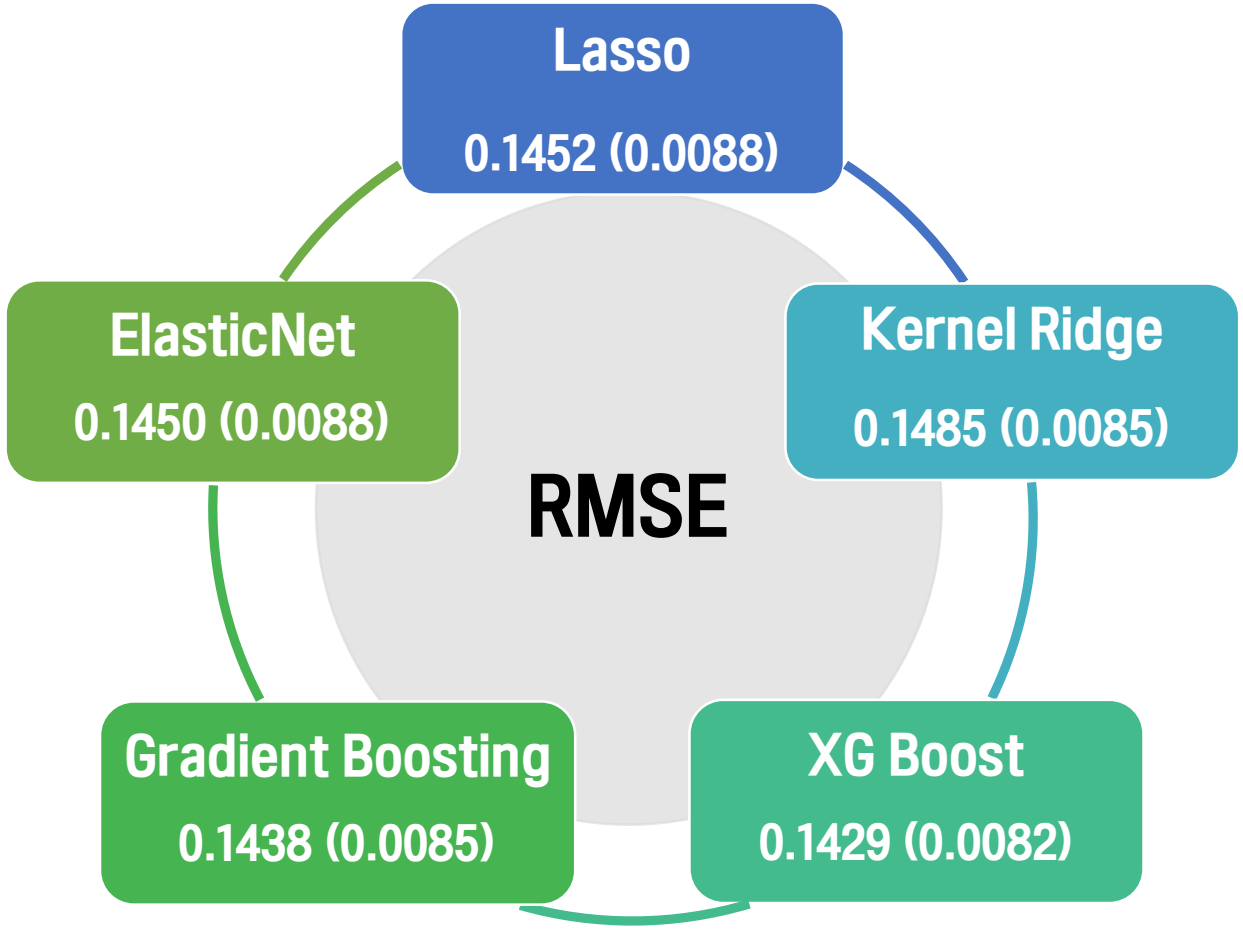
- After Normalization



2 Trial and Error

1) First Try

Cross Validation



2 Trial and Error

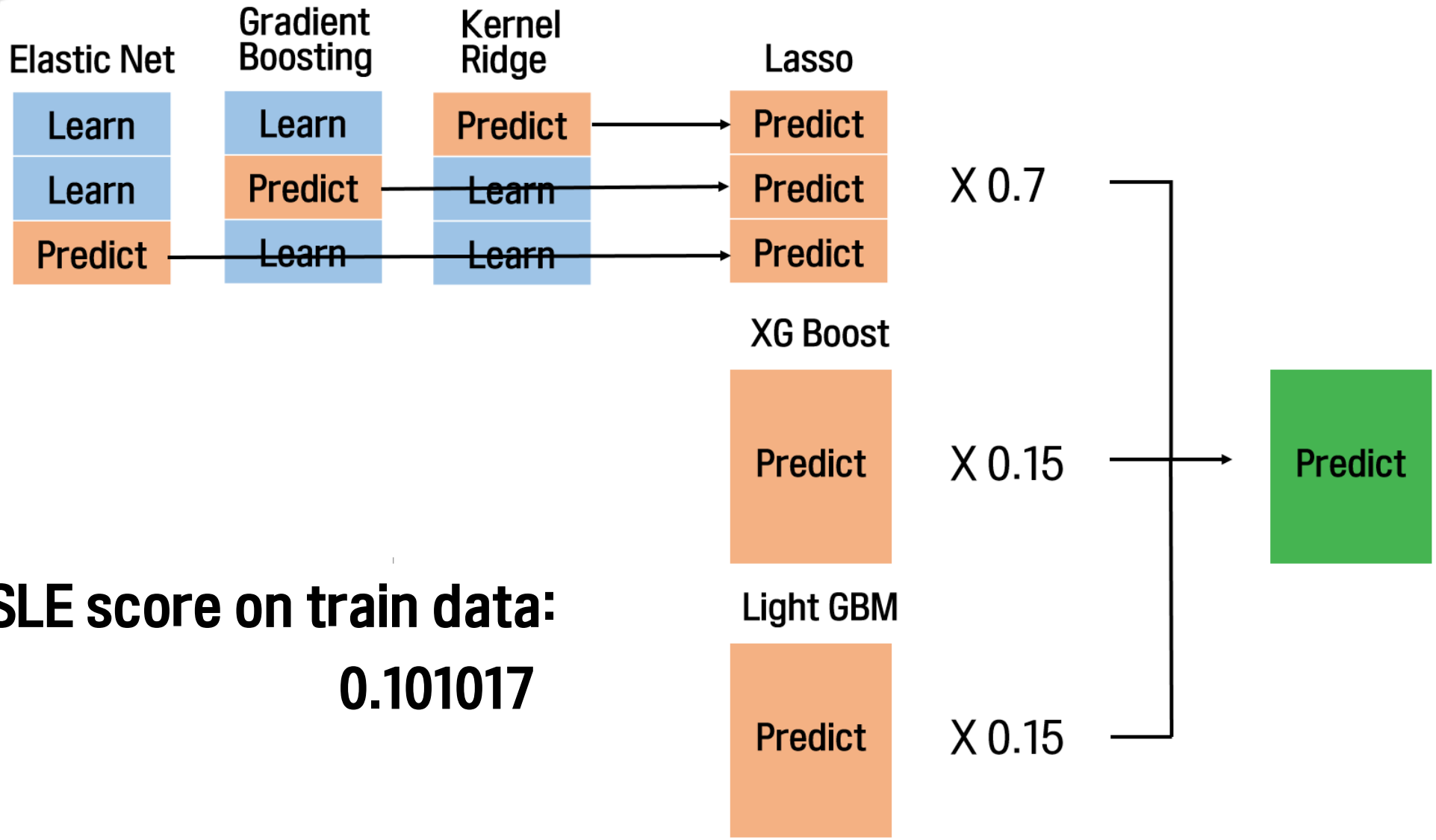
2) Second Try

- Related work : A Hybrid Regression Technique for House Prices Prediction

Features	Hybrid Method	Score
230	0.65Ridge+0.35Xgb	0.11318
230	0.70Lasso+0.30Xgb	0.11294
230	0.65Lasso+0.35Xgb	**0.11260**
230	0.60Lasso+0.40Xgb	0.11277
230	0.3Ridge+0.35Lasso+0.35Xgb	0.11285
230	0.25Ridge+0.40Lasso+0.35Xgb	0.11283
280	0.65Ridge+0.35Xgb	0.11458
280	0.65Lasso+0.35Xgb	0.11539

Tab. 5 Hybrid Combination of Regressions

2 Trial and Error

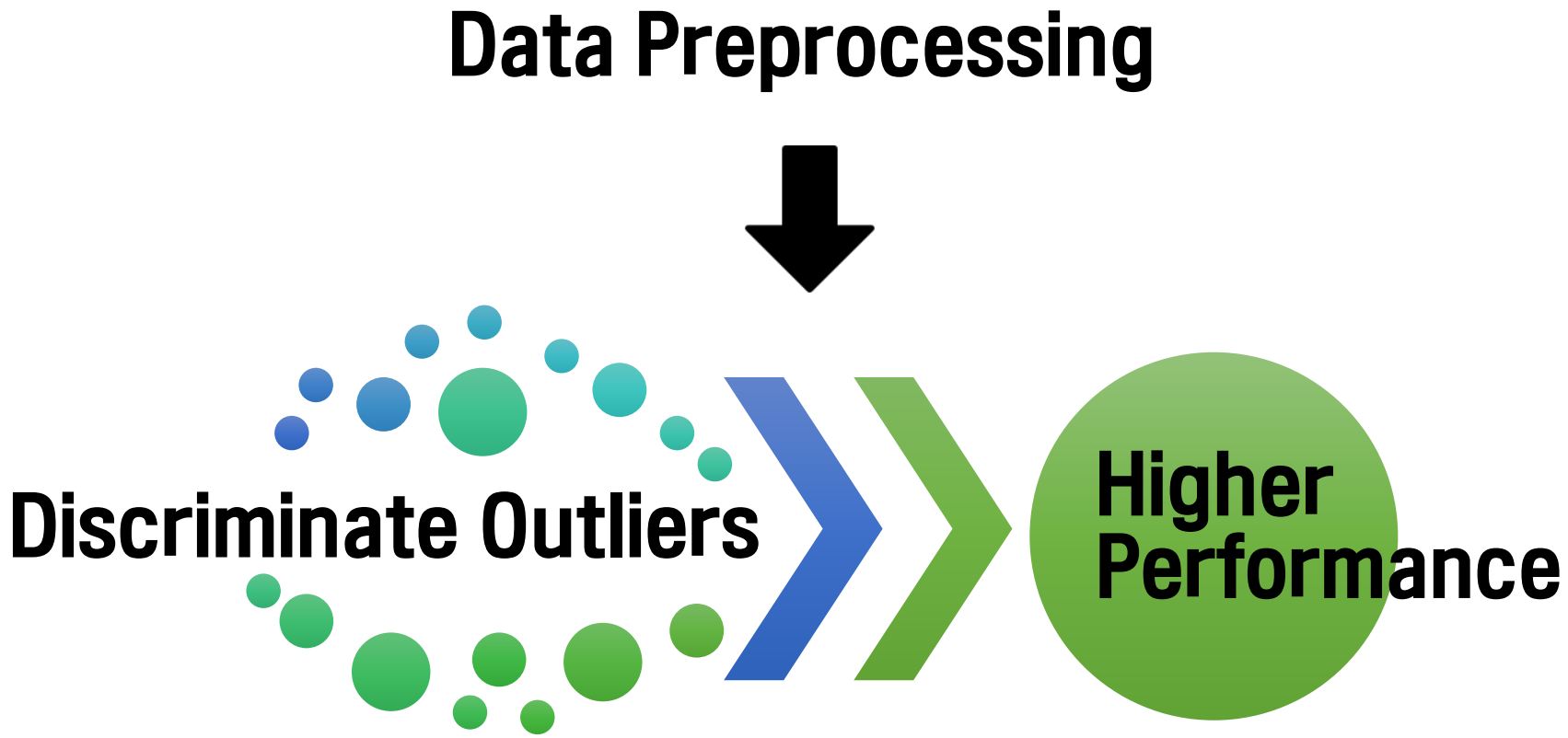


RMSLE score on train data:

0.101017

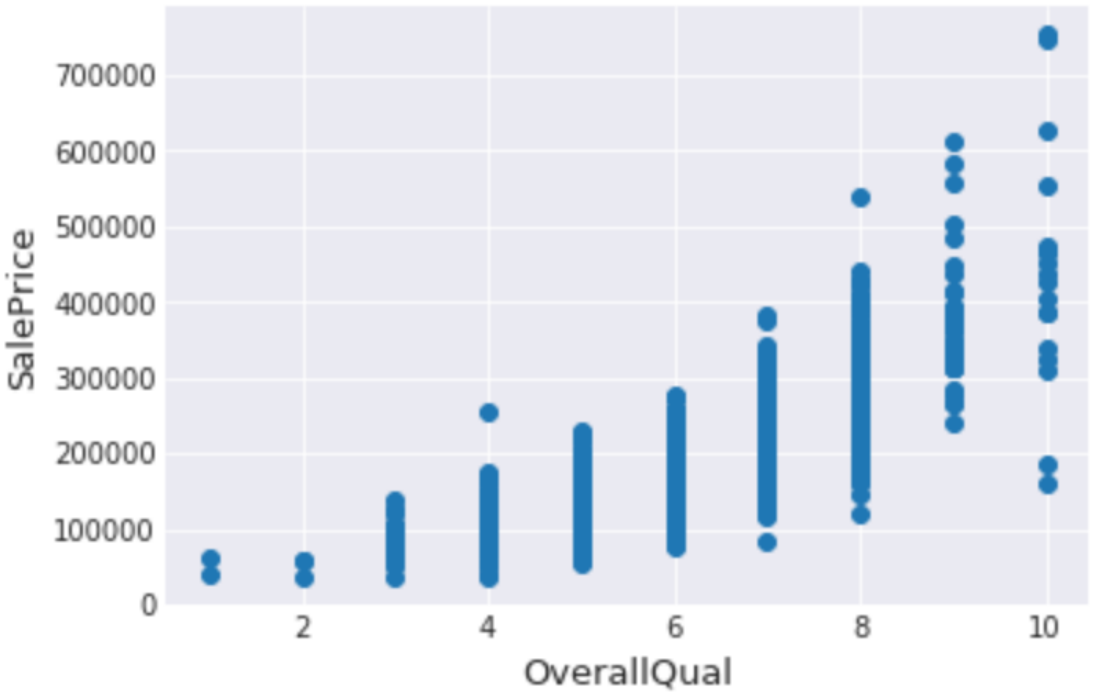
2 Trial and Error

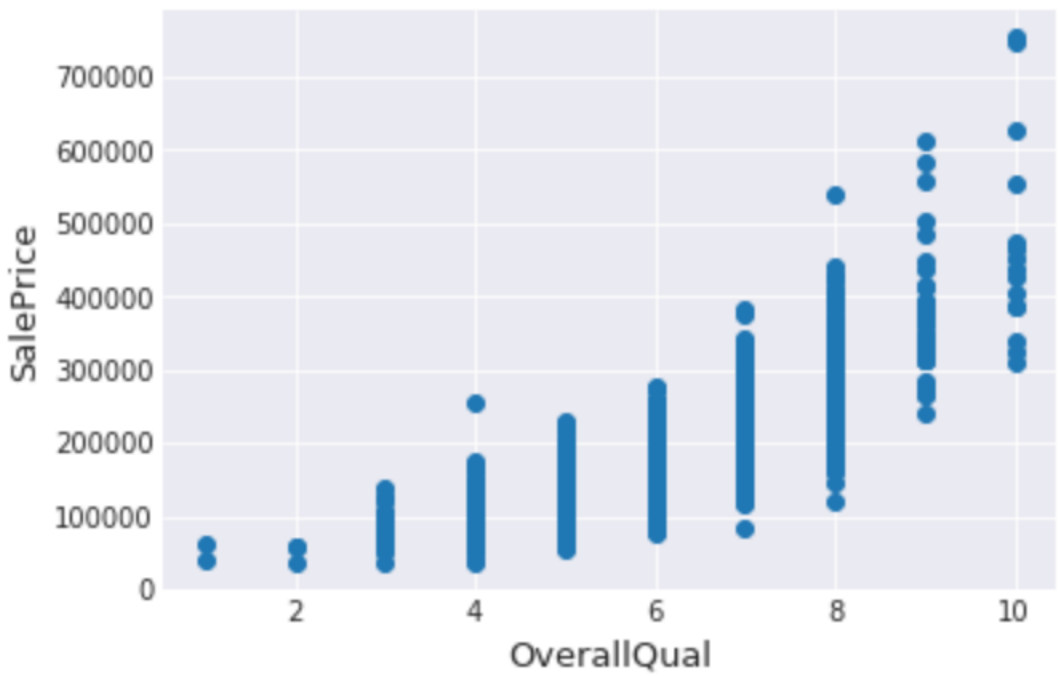
3) Third Try



2 Trial and Error

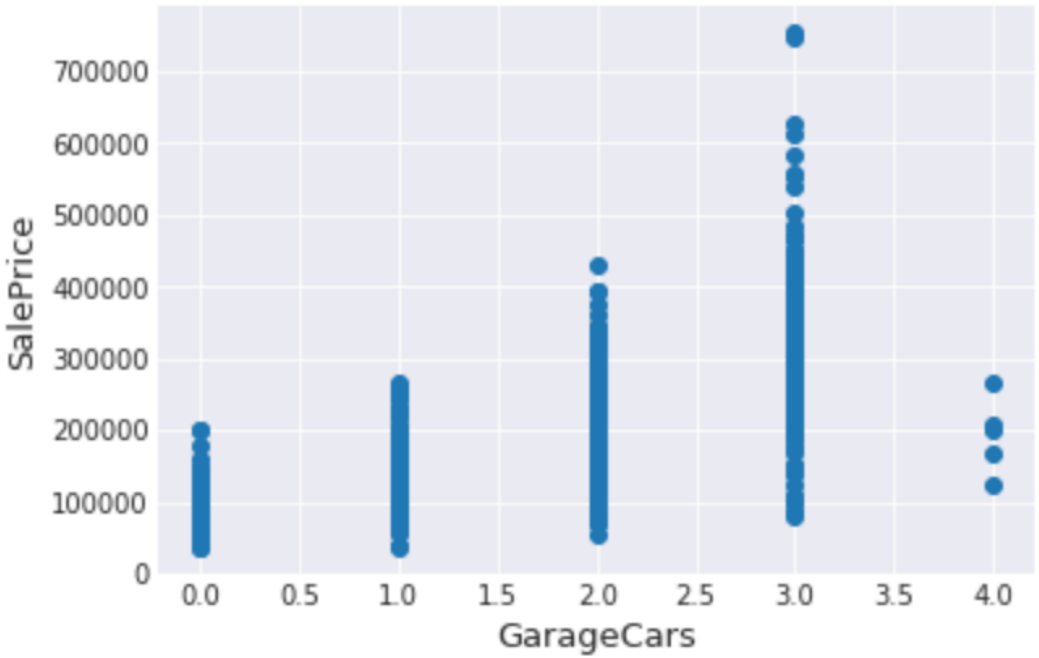
3) Third Try

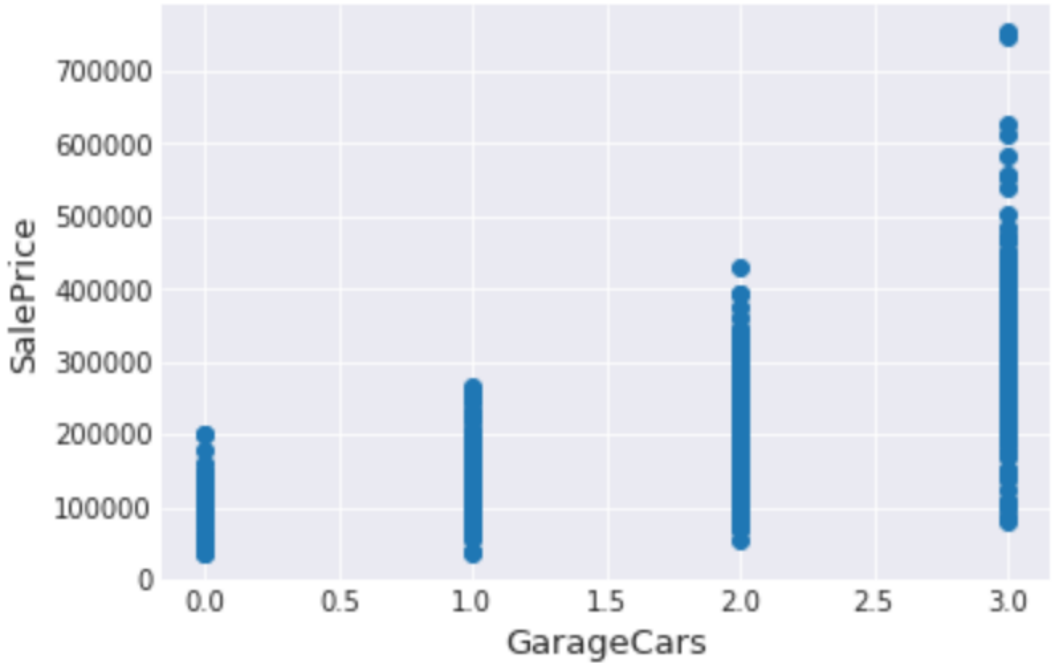




2 Trial and Error

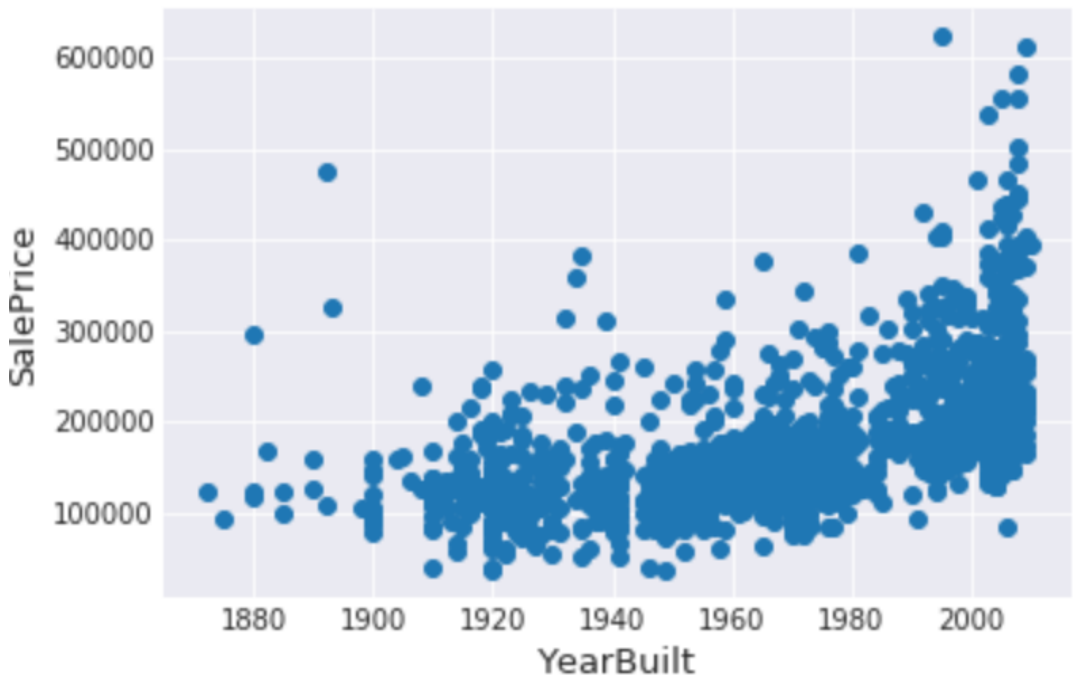
3) Third Try

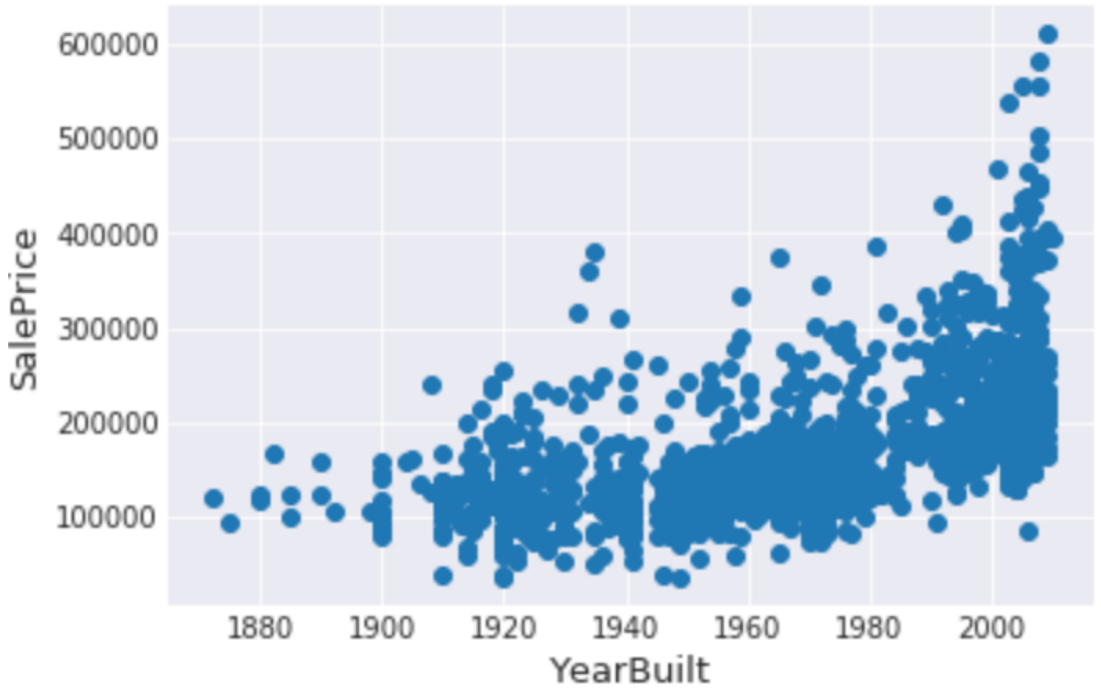




2 Trial and Error

3) Third Try

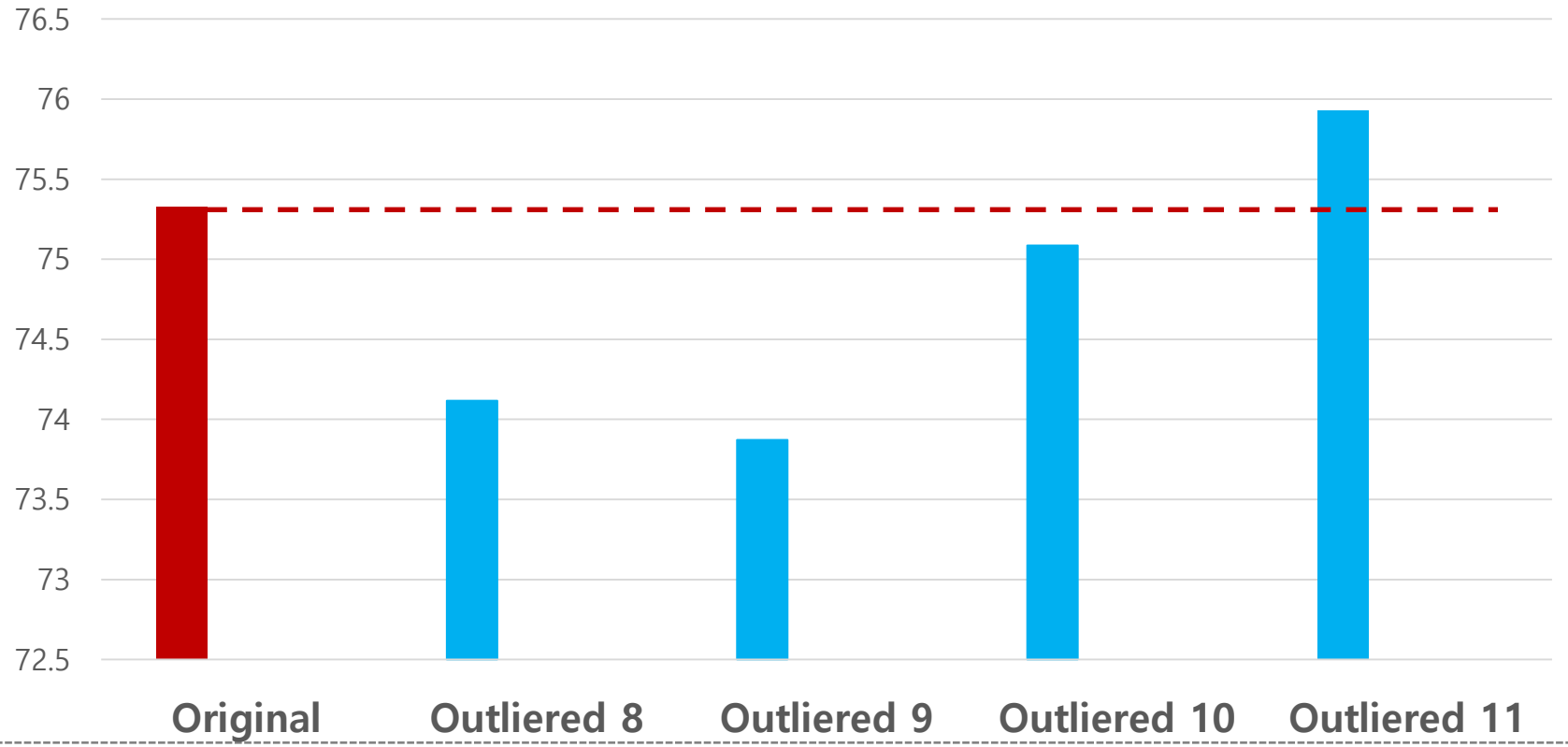




2 Trial and Error

3) Third Try

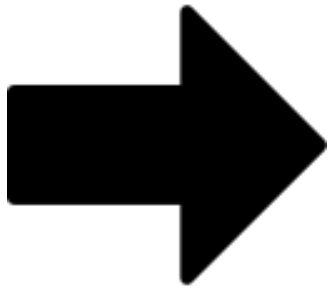
Performance based on the Number of Outliered Features



3 Conclusion

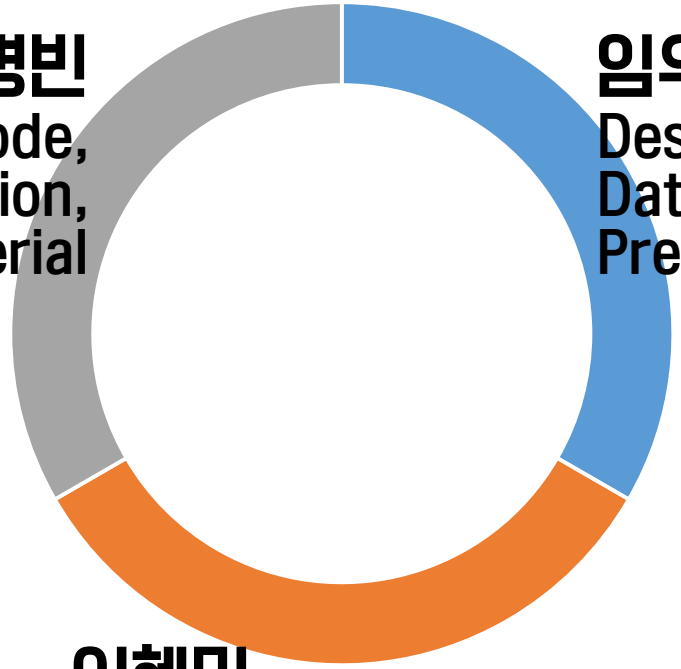
Total Rank in Kaggle

- First Try:
- Second Try: 593
- Third Try: 220



0.4%

4 Team Participation & References



조경빈
Design, Code,
Data Analyze, Data Collection,
Presentation Material

임우담
Design, Code, Data Analyze,
Data Collection,
Presentation

이혜민
Design, Code,
Data Analyze, Presentation,
Presentation Material

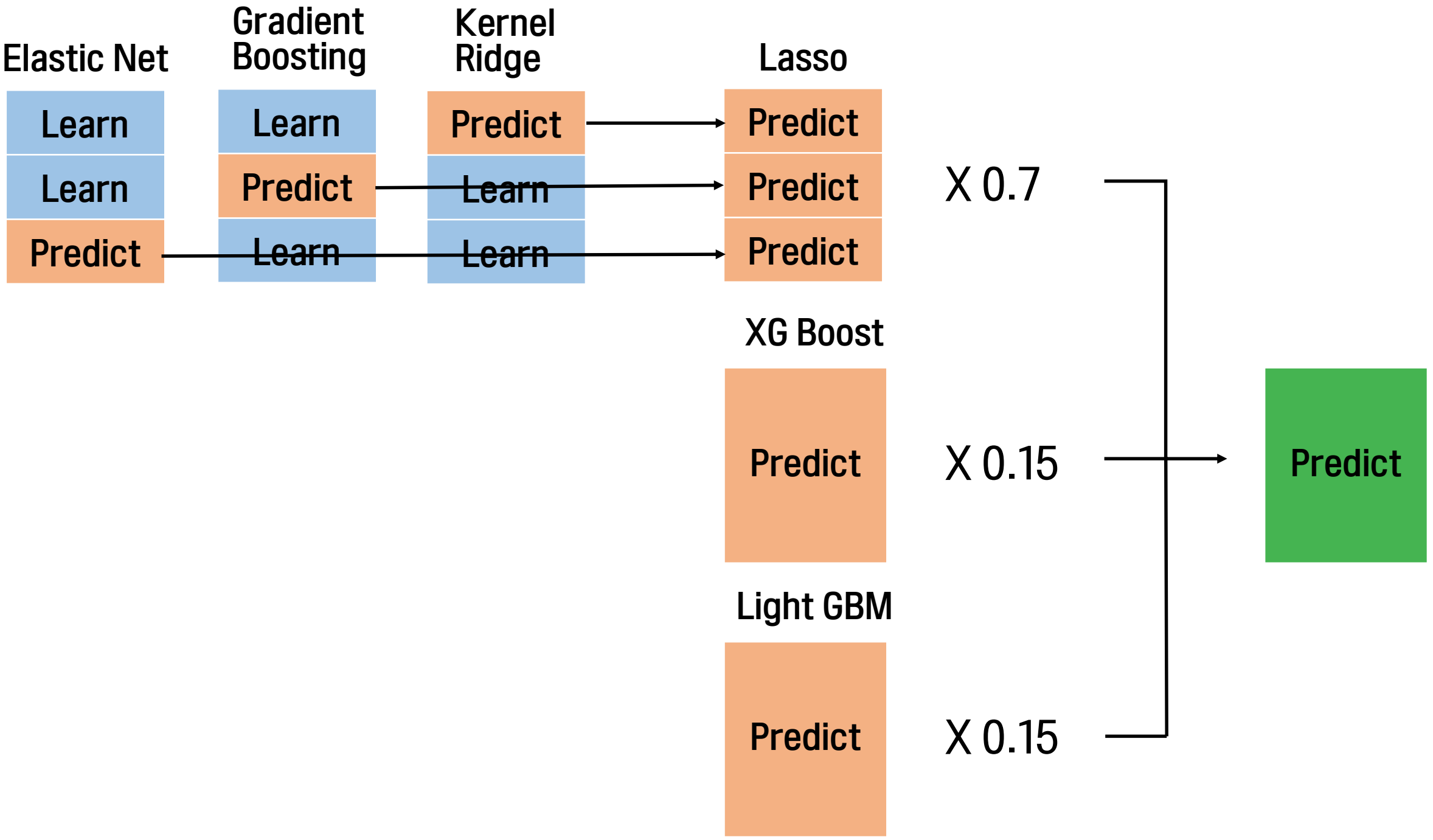
4 Team Participation & References

- Adyan Nur Alfiyatin, Hilman Taufiq, Wayan Firdaus Mahmudy, Ruth Ema Febrita, 「Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization」, International Journal of Advanced Computer Science and Applications, 2017

- Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong Goh, 「A Hybrid Regression Technique for House Prices Prediction」, 2017

Thank You

AI Project: House Price Prediction


Elastic Net
Learn
Learn
Predict
Gradient Boosting
Learn
Predict
Learn
Kernel Ridge
Predict
Learn
Learn
Lasso
Predict
Predict
Predict
X 0.7
XG Boost
Predict
X 0.15
Light GBM
Predict
X 0.15
Predict