1 Analysis

A number of features are excluded before training the model. The exclusion was determined by investigating the pairwise correlation between the complete set of features and the valuation price. A list of the excluded features are found in appendix 11.0.1. The model used for ML are found in appendix 11.0.2.

The initial 5-fold cross-validation to obtain the optimal values of the hyperparameters in each regularization yields, can be seen in table 1.

Table 1: Optimal hyperparameters, two degrees polynomial features

	λ_{Lasso}	λ_{Ridge}	$\lambda_{ElasticNet}$
2 degrees	372.76	26.83	1
3 degrees	2310.13	432.88	1

The prediction-errors of each model, with two and three degrees of polynomial features respectively, are printed below:

Table 2: Prediction Errors, two degrees of polynomial features

	Lasso	Ridge	Elastic Net	OLS
MSE	59,081,597,331.11	62,416,553,466.45	1	64,297,245,726.80
RMSE	243,067.06	249,833.05	1	253,569.02
MAE	47,660.52	51,306.19	1	52,035.52

Table 3: Prediction Errors, three degrees of polynomial features

	Lasso	Ridge	Elastic Net	OLS
MSE	60,103,955,188.37	55,667,264,851.14	85,824,334,267.16	2.848e+28
RMSE	245,161.08	235,939.11	292,957.91	168,763,945,674,352.3
MAE	47,126.43	43,630.55	69,582.56	2,637,476,978,906.39

Since our efforts are aimed at predicting the valuation of a given real-estate, we want to penalize the size of the error the MSE-score is a better indicator for our model's performance as opposed to the mean absolute error (MAE). Since MSE squares the error terms, it gives a relatively high weight to large errors, which is desirable for our current aim.

From table XX it is evident that the machine learning model with two degrees of polynomial features performs best with a regularization by LASSO as it has the lowest measures of difference between the predicted price and the true price, with a MSE of 5.91E+10, whilst the

model with 3 degrees of polynomial features regularized by ridge regression performs even better with a MSE 5.57E+10. Therefore, we chose this model as the final model from which we will conduct the rest of the analysis with.

2 Results

Figure 1: Predicted Valuation vs. Actual Valuation