Linear Data Housing Appraiser (LinDHA)

Lost $500 Million in 2021

“Good Idea, Bad Execution”

Machine Learning as a helper,
NOT as a replacer

Idea:

1) Build a simple model to estimate Sale Prices and identify opportunities.

2) Send in a human expert for proper appraisal (Zillow skipped this step!)

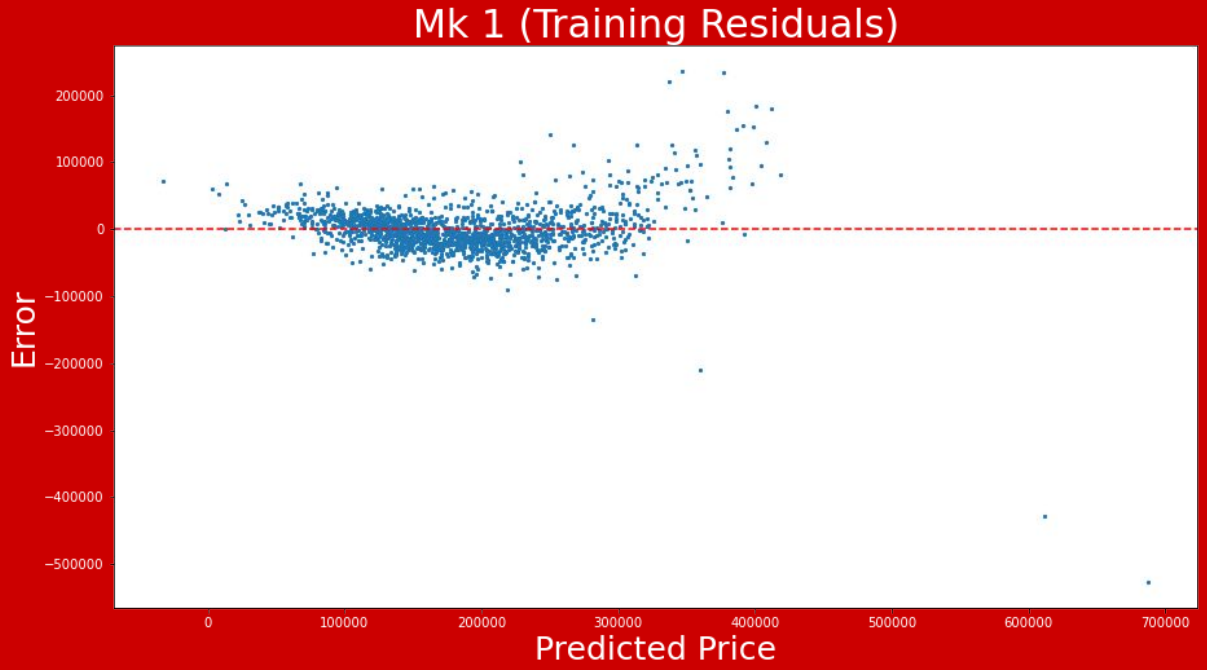
3) Lease/License it to real estate industry / banking+mortgage industry.

4) Use customer generated data to live service and update model.

Objective:

Create a linear model with the lowest Possible Mean Absolute Error (MAE)

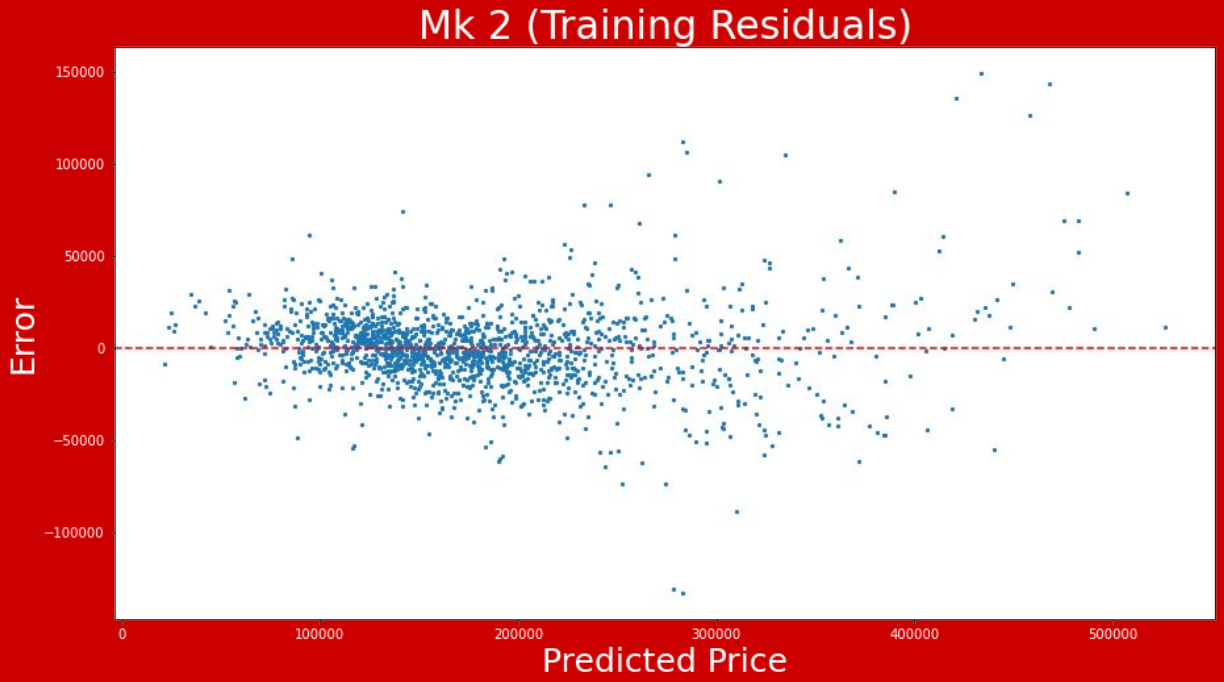
LinDHA Mk. 1 (Base Model)



- 12 numerical features
  - Underfitting (bias)
  - Outlier: House 2181
- Test MAE = $21,400

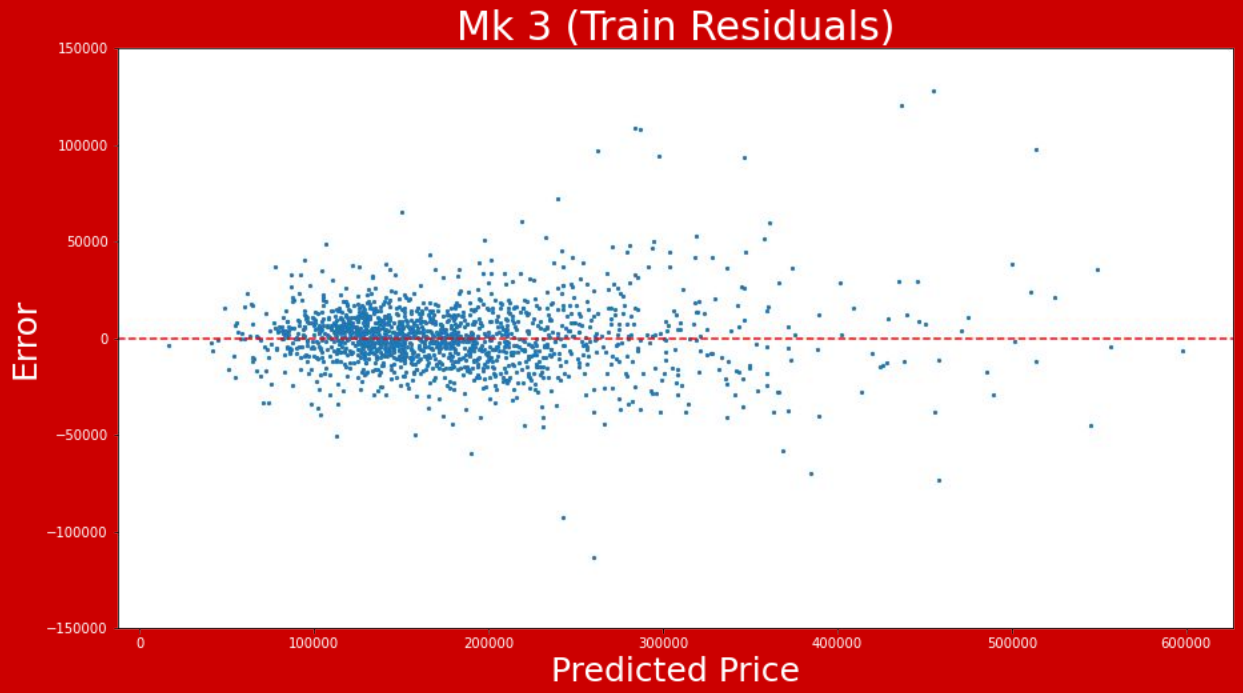
STARK INDUSTRIES
BETTER LIVING THROUGH TECHNOLOGY

LinDHA Mk. 2 (Feature Selection)



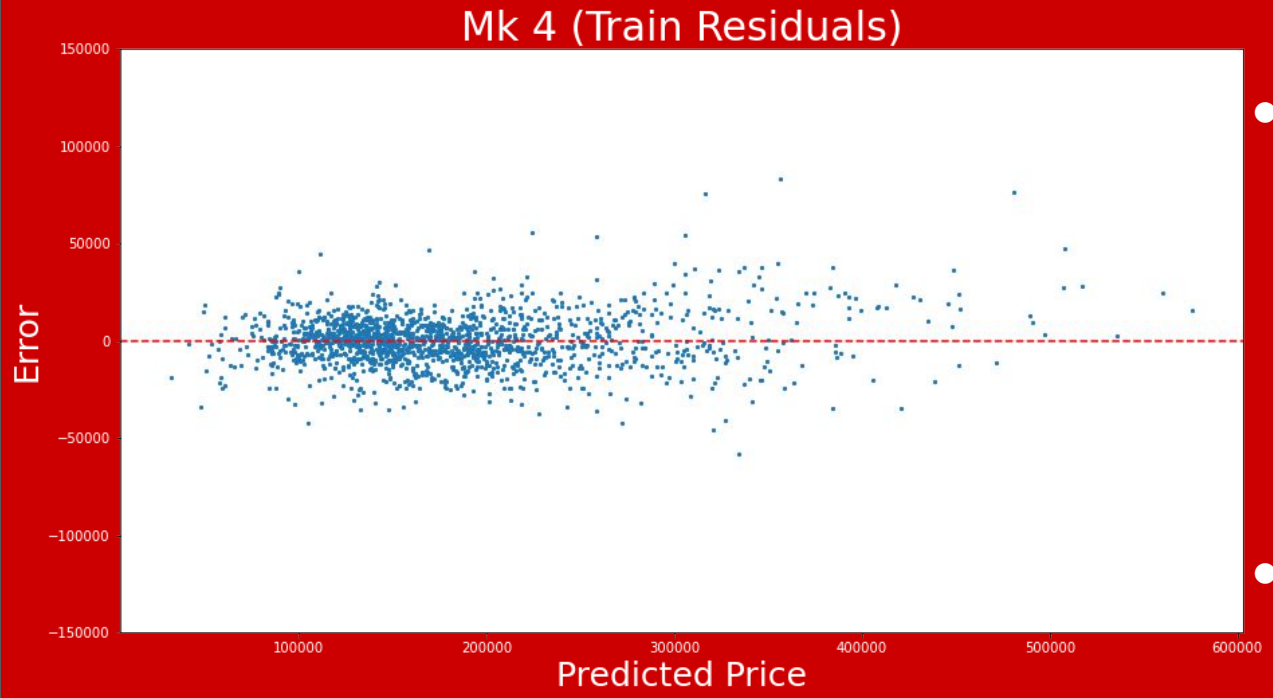
- 40 Features
  - 22 Numerical
  - 18 Categorical
  - Slight underfitting
- Best Test MAE = $15,700

LinDHA Mk. 3 (Feature Engineering)



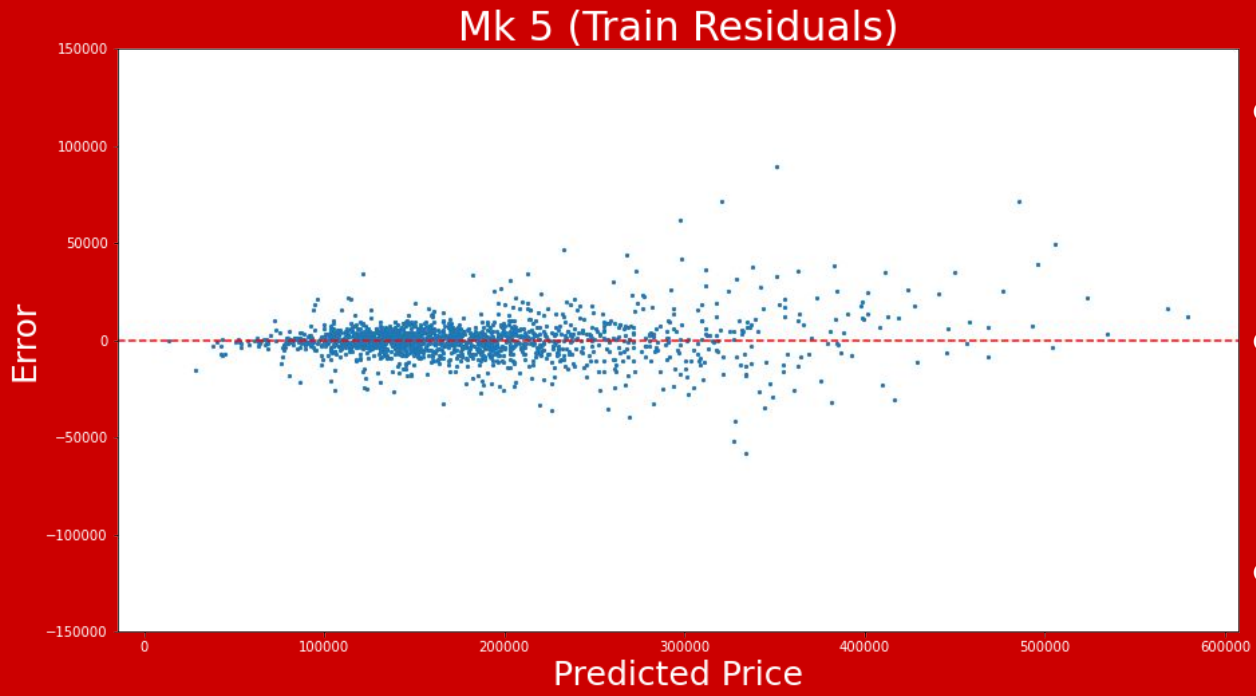
- 62 Features
  - 44 Numerical
  - 18 Categorical
  - Interaction Terms
  - Heteroscedastic residuals
- Best Test MAE = $13,800

LinDHA Mk. 4 (LASSO)



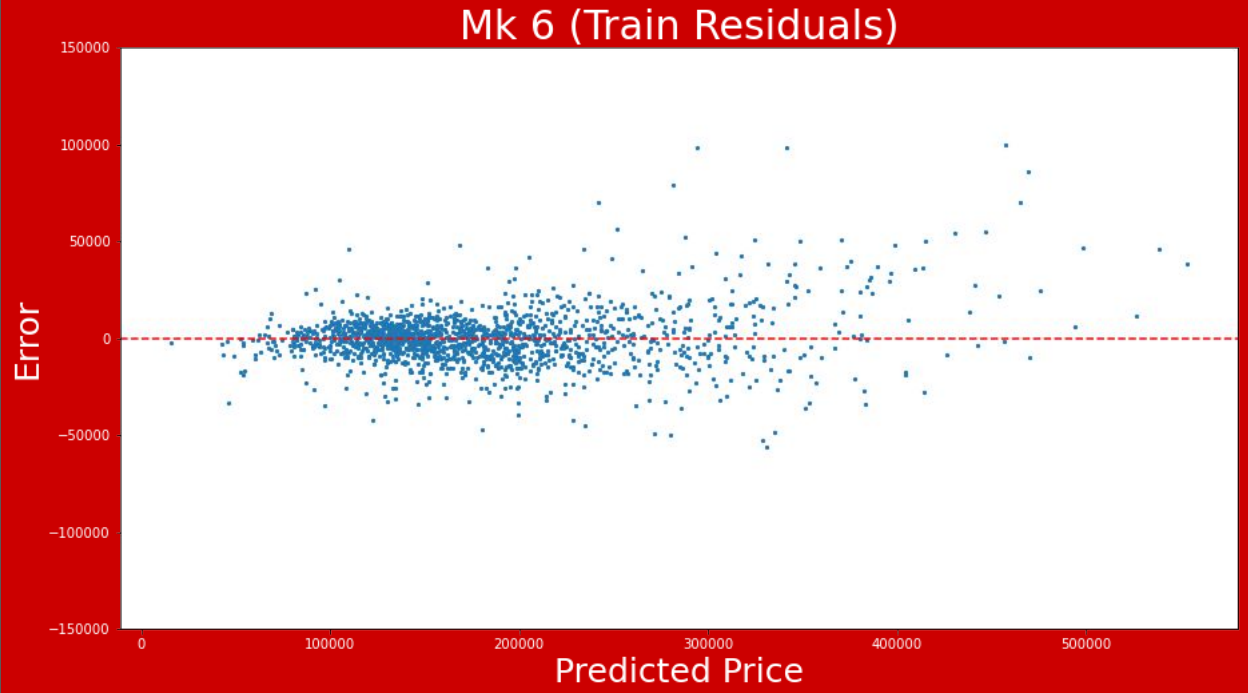
- Started with 42,000+ features
  - Lasso-ed down to 522
  - Residuals well-behaved
- Best Test MAE = $13,600

LinDHA Mk. 5 (Ridge)



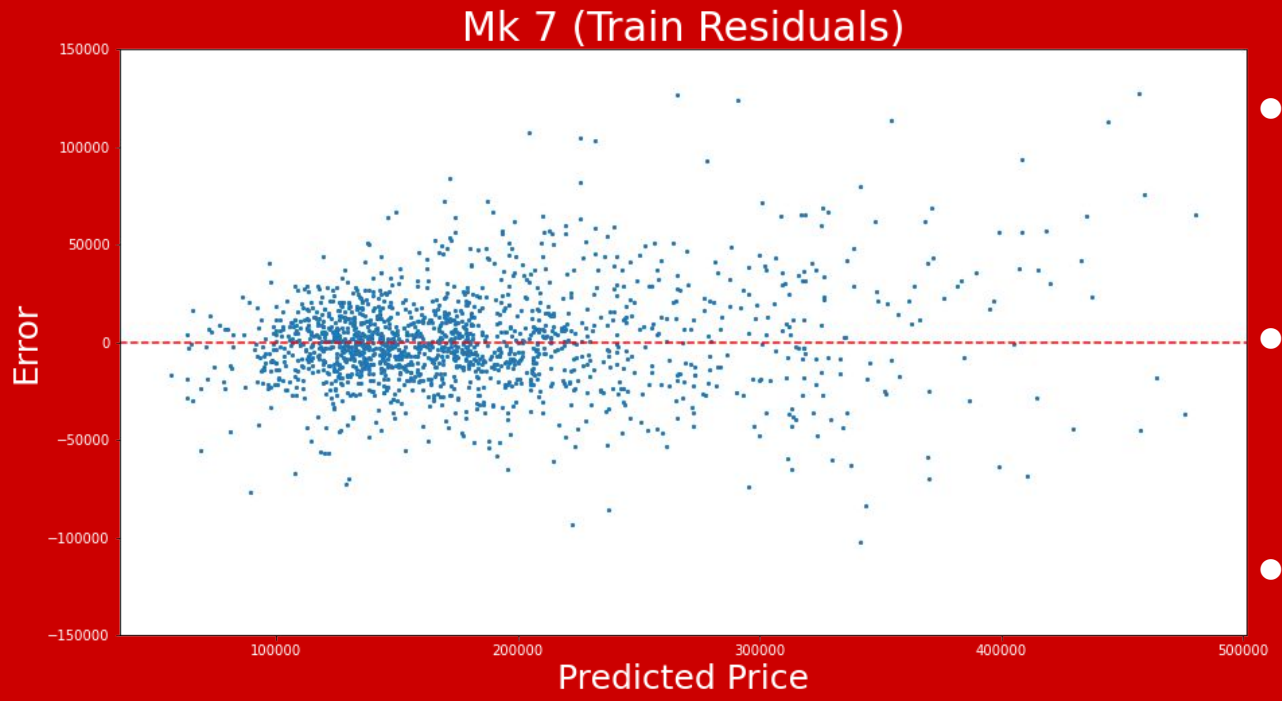
- 42,000+ Features.
  - Ridge reduced to 25,000+ features
- Tried 299 Features
  - Even worse performance
- Best Test MAE = $15,400

LinDHA Mk. 6 (Elastic Net)



- Paused grid search at alpha=26, L1_ratio = 0.65
- Model was still improving, but search time was becoming too costly
- Best Test MAE = $15,300

NeaNHA Mk. 1 (KNN)



- Nearest Neighbors = 10 (Manhattan/Taxicab Norm)
- Expensive homes no longer underestimated!
- Best Test MAE = $20,800

Summary and Recommendations

- ★ On average: LinDHA predicts within $14,000 to $16,000.
  - Use LinDHA to “detect” good/bad home investments.
  - DO NOT use LinDHA to “replace” human appraisers.
- ★ Can use LinDHA to detect “anomalies” like House 2181.
  - Might be viable product in mortgage-lending and financing sector.
- ★ Room for Improvement:
  - Elastic Net can be tuned (probably the best model to use moving forward).
  - NeaNHa worth further testing (curse of dimensionality is a problem though).