Maximizing profits with ML homeflipping strategies.

by: Laurel He, Daniel Setiawan, Daniel Erickson

1/27/23

Introduction

Home flipping is economically sexy... ***or is it?***

- ➔ 20% profit is the *break-even* point.
- ➔ 28% of home flippers **lose** money.

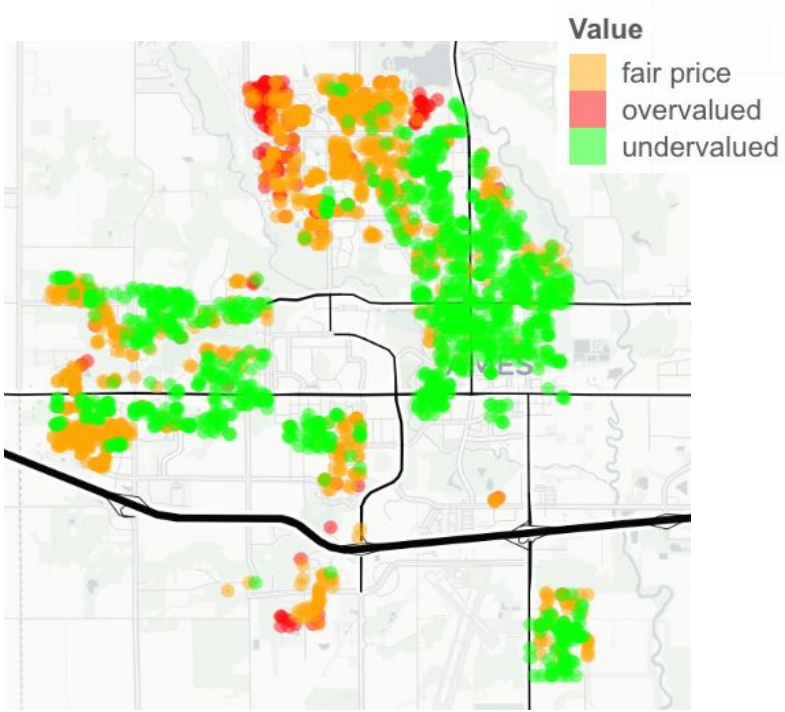
Successful flips require thorough research & ***in-depth analysis*** of the housing market.

Throughout this presentation, we'll use **ML techniques** to make fiscally smart decisions.

Let’s talk data.

Ames, IA

- 28 neighborhoods
- 4 years (2006 - 2010)
- 78 features
- 2500 observations



But the data is so dirty... what to do?

Preprocessing: Cleaning the data

Data Selection

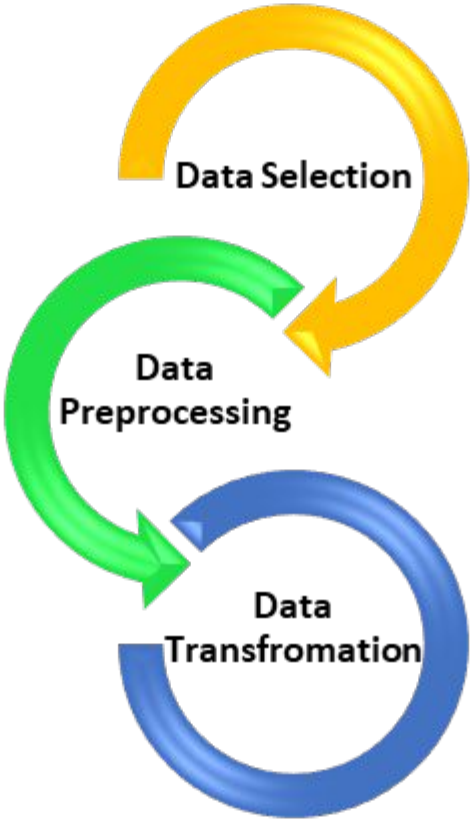
- Removed collinear features (for linear models)
- Removed extremes / outliers in target (SalePrice)

Data Preprocessing

- Different models require different preprocessing
  - 3 separate data variants

Data Transformation

- Standardization: $X' = \frac{X - \mu}{\sigma}$



Feature Engineering: Cleaning the data

Bringing in New Info:

Crime Rates : 0 **(safe)** → 10 **(unsafe)**

Public Schools : 0 **(worst)** → 10 **(best)**

Real Estate Appreciation : 0 **(lowest)** → 10 **(highest)**

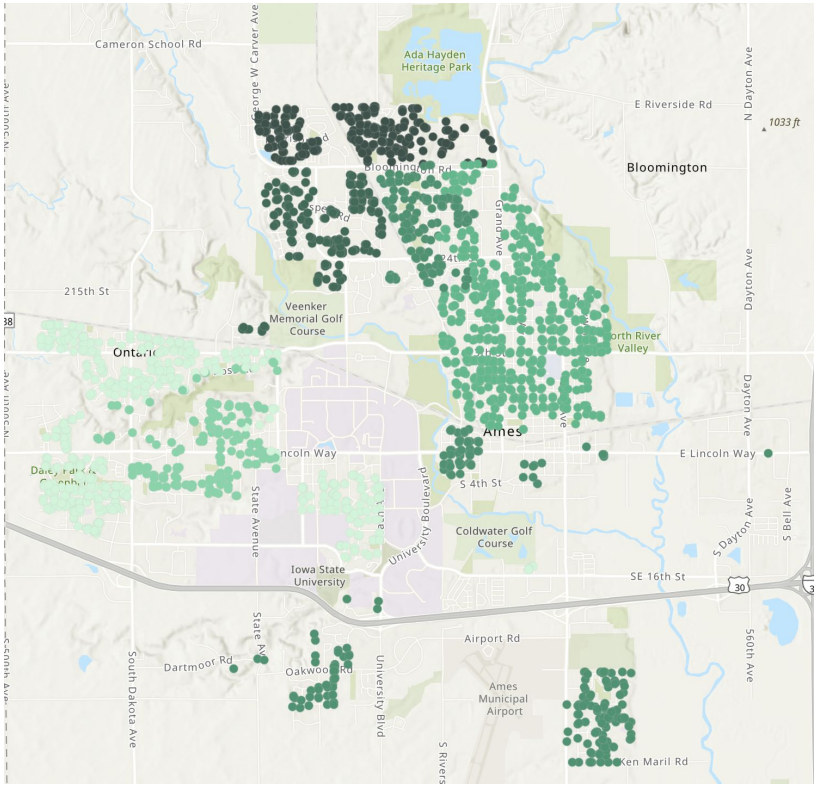
Income Growth Rate : 0 **(min)** → 5 **(no change)** → 10 **(max)**

Reduce feature multicollinearity:

HouseAge : YrSold - YearBuilt

LastRemod : YrSold - YearRemodAdd

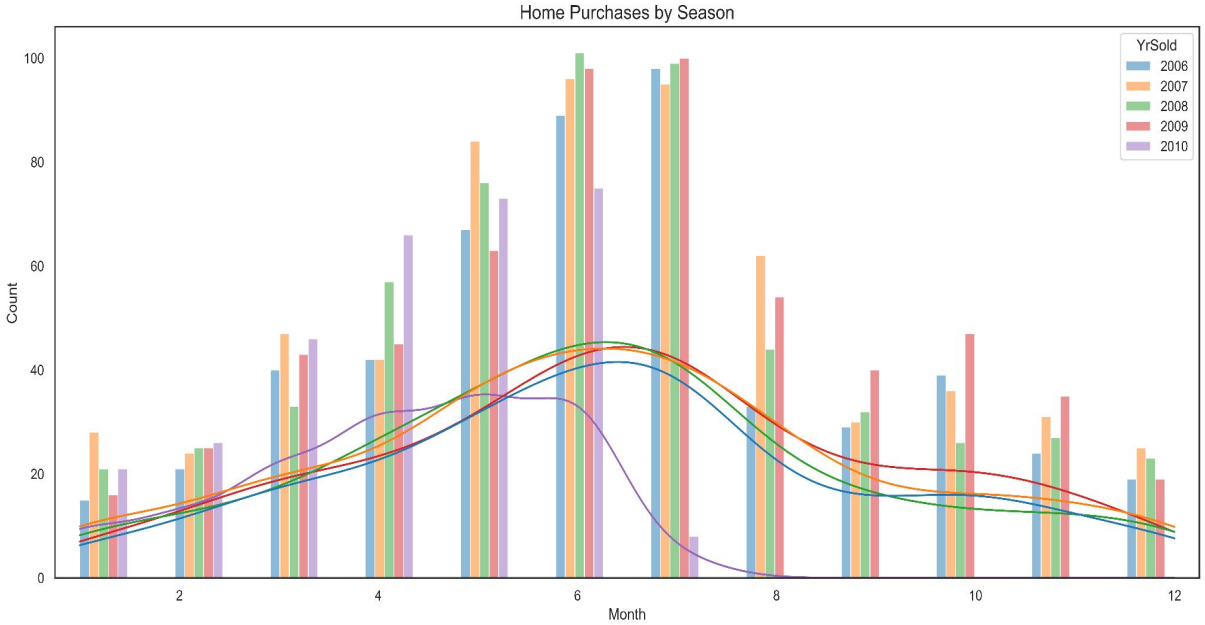
Public School Ratings in Ames



Reference: https://www.neighborhoodscout.com/ia/ames

The **when** & **where**.

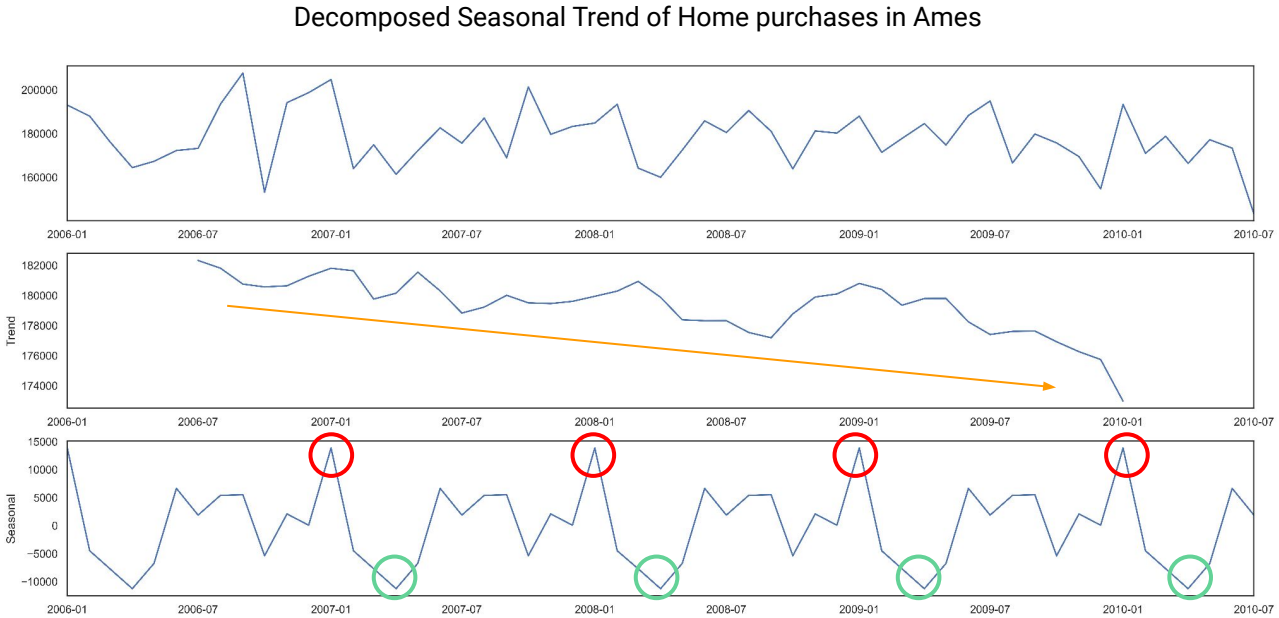
Discovery: Tis the season to be wealthy.



Seasonality directly impacts supply & demand.

Summer sees an influx of demand, increasing costs.

Decomposition: Time is a series of fluctuating variables.

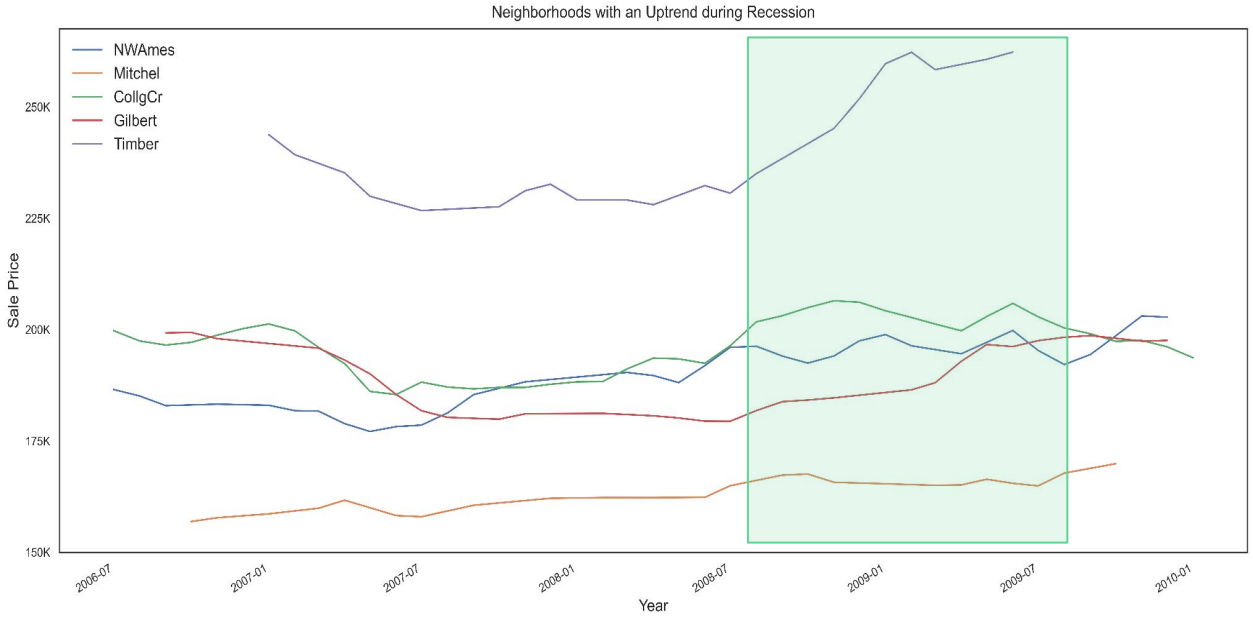


* Trend also plays a significant factor to the prices of homes

Decomposing trend & seasonality provides clarity.

Best to buy: **March**
Best to sell: **January**

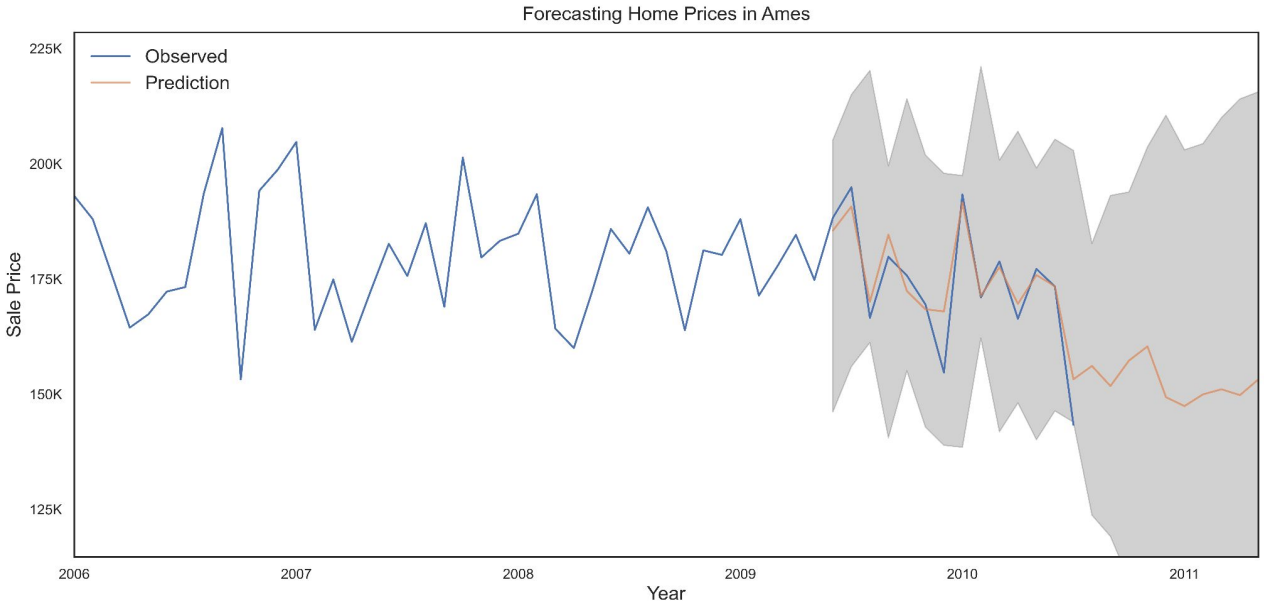
Resiliency: Don’t buy the house; buy the neighborhood.



Some neighborhoods are more ***recession-proof***

Premium neighborhoods for recession 2023

Forecasting: I see, with my ~~crystal ball~~ charts and graphs.

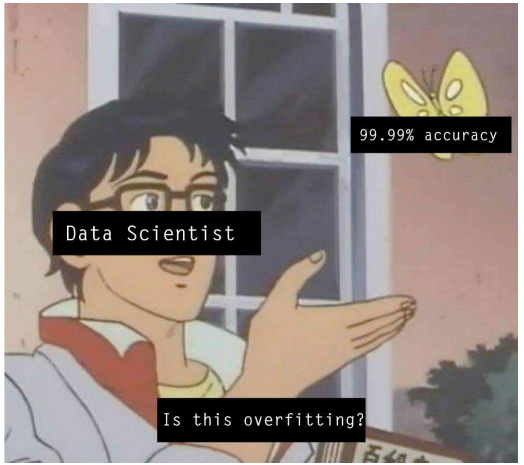


Adjusting for **differencing**, **seasonality** & **trend** can help forecast future prices and predict optimal timing to buy / sell more efficiently.

The **what**.

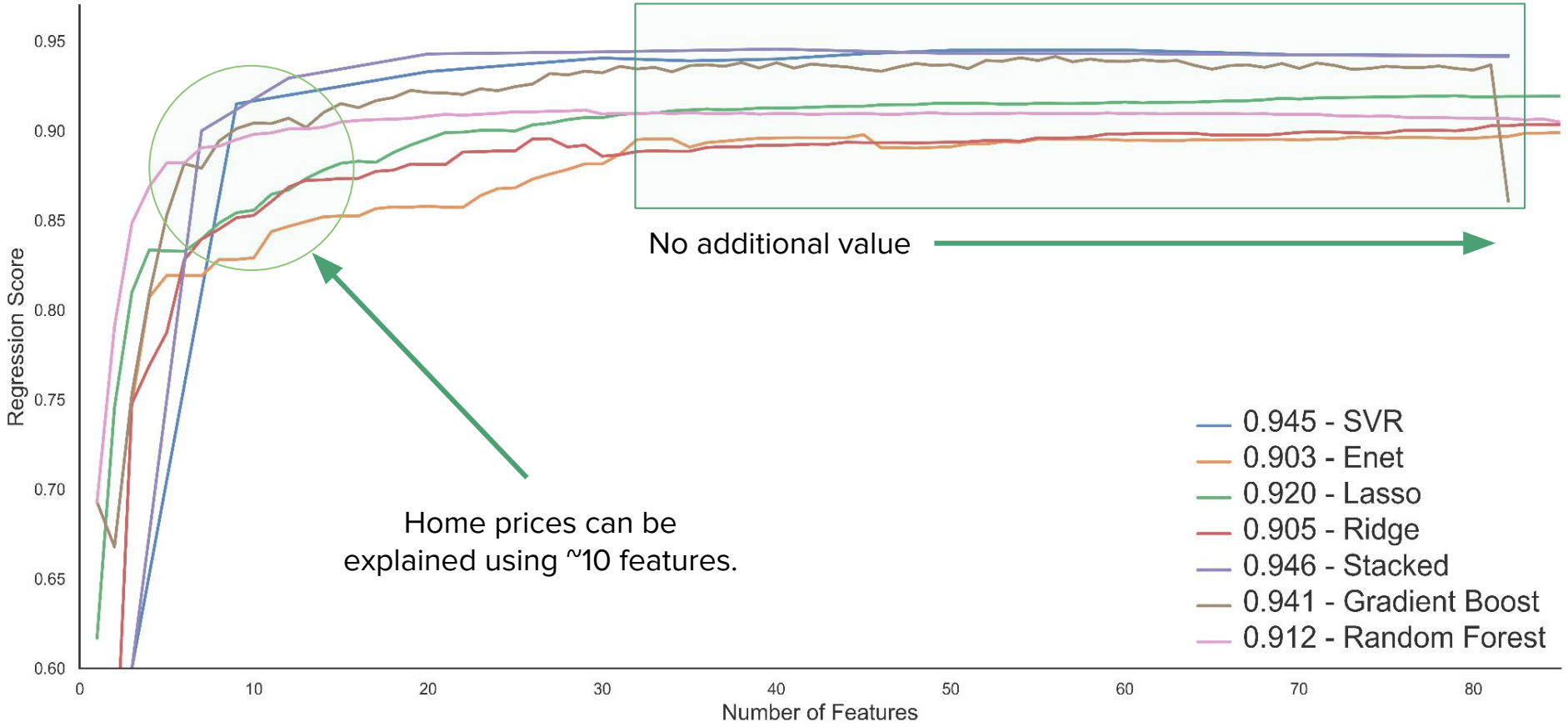
Model Scores: Summarizing the summary of the summary.

Model	Cross-Validated R^2 (test set)	Hyperparameters	Preprocessing
Lasso	0.920 (CV = 10)	α : 1.00	**1.** Removed Collinearity **2.** Dummified nominal **2.** Standard Scaled
Ridge	0.905 (CV = 10)	α : 1.00	
Elastic Net	0.903 (CV = 10)	α : 1e-5	
SVR	0.945 (CV = 5)	C : 10, ε : 0.04, γ : 0.002	**1.** Label Encoding **2.** Standard Scaled
Random Forest	0.912 (CV = 5)	Trees : 500 Depth : 10 Sample Leafs : 2 Sample Split : 5	**1.** Label Encoding
Gradient Boosting	0.941 (CV = 3)	Trees : 10K Learning Rate : 0.1 Depth : 3 Sample Leafs : 1 Sample Split : 2	**1.** Label Encoding **2.** Standard Scaled
Stacked Model	0.946 (CV = 5)	GB, SVR and RF	**1.** Label Encoding **2.** Standard Scaled

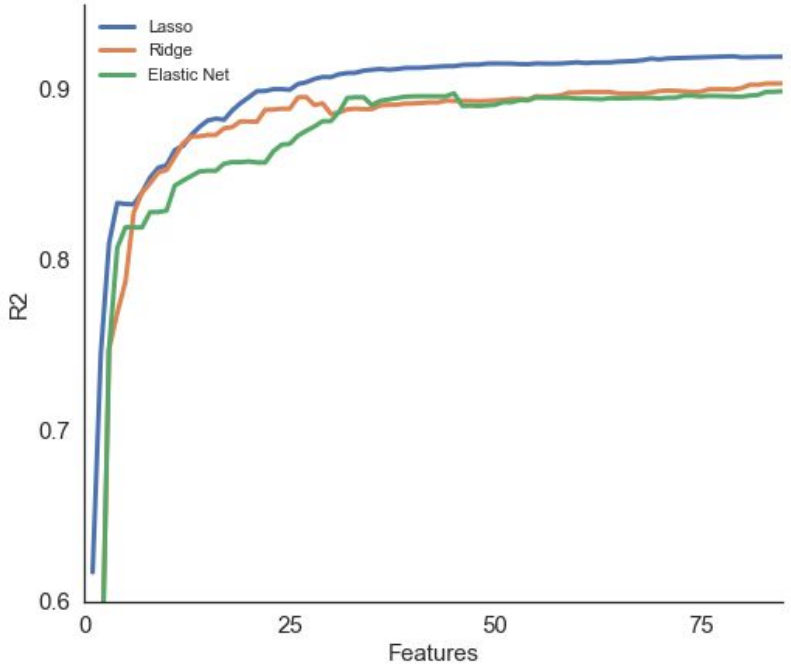


* App uses Lasso model

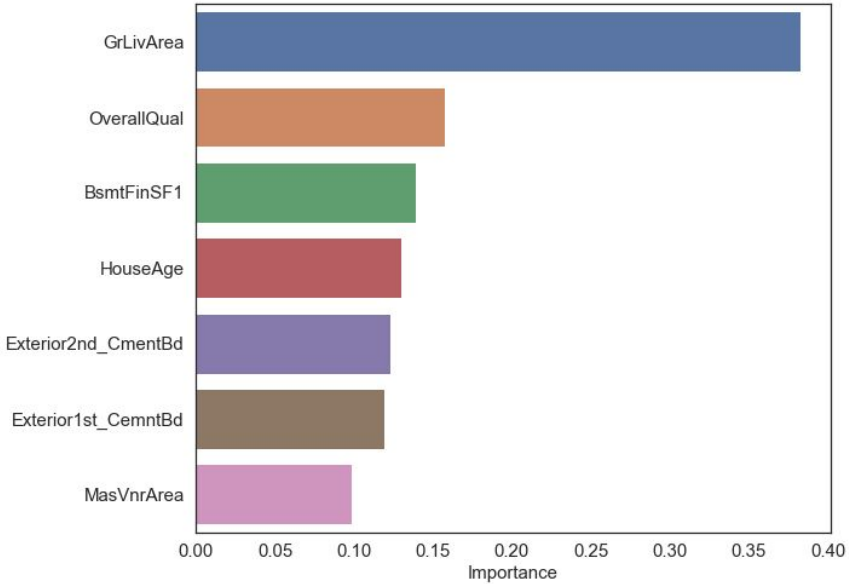
Model Performance v. Features



Penalized Regression: Lasso, Ridge, Elastic Net.



Top Important Features, Lasso:

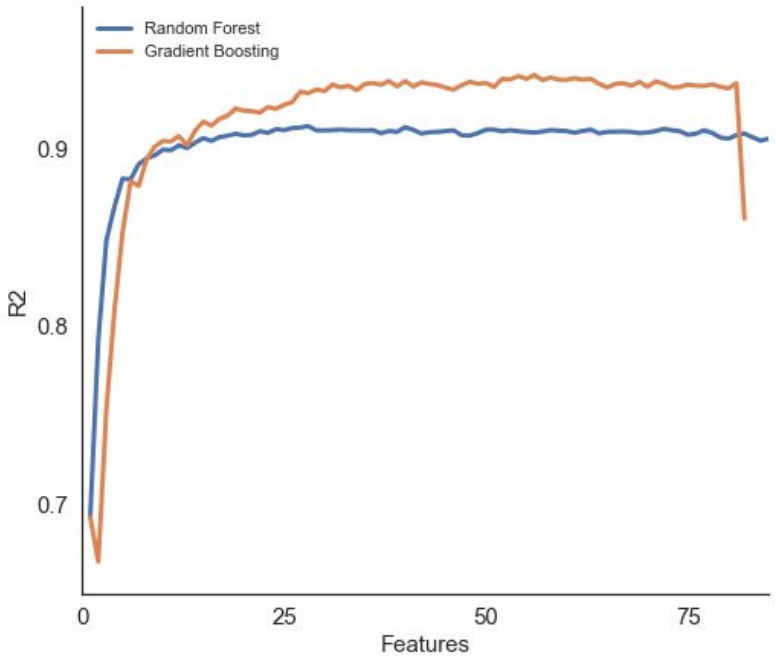


Lasso:
α : 1.00

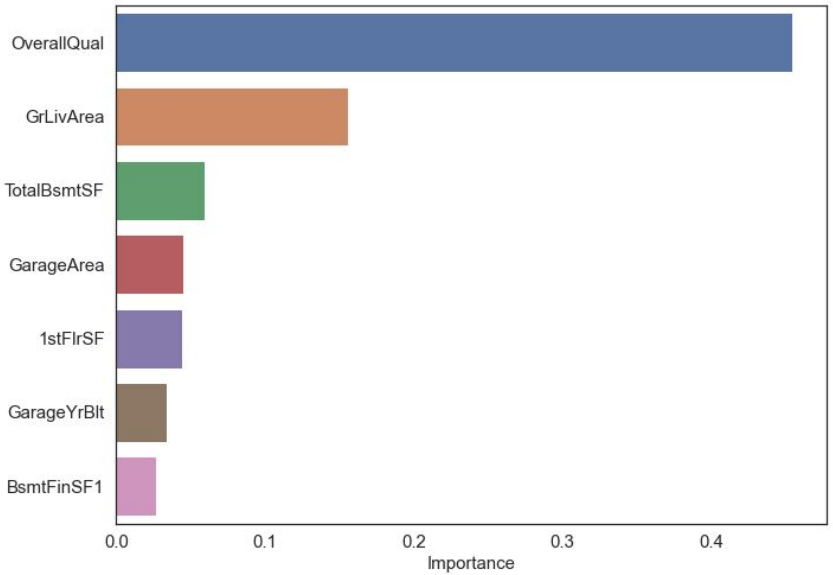
Ridge:
α : 1.00

Enet:
α : 1e-5

Trees: Random Forest, Gradient Boosting



Top Important Features, Gradient Boost:



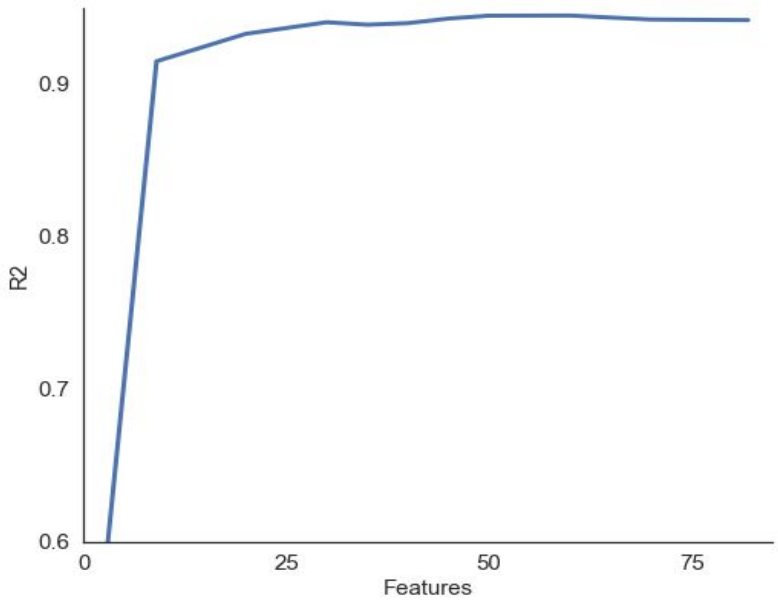
RF:

Trees : 500
Depth : 10
Sample Leafs : 2
Sample Split : 5

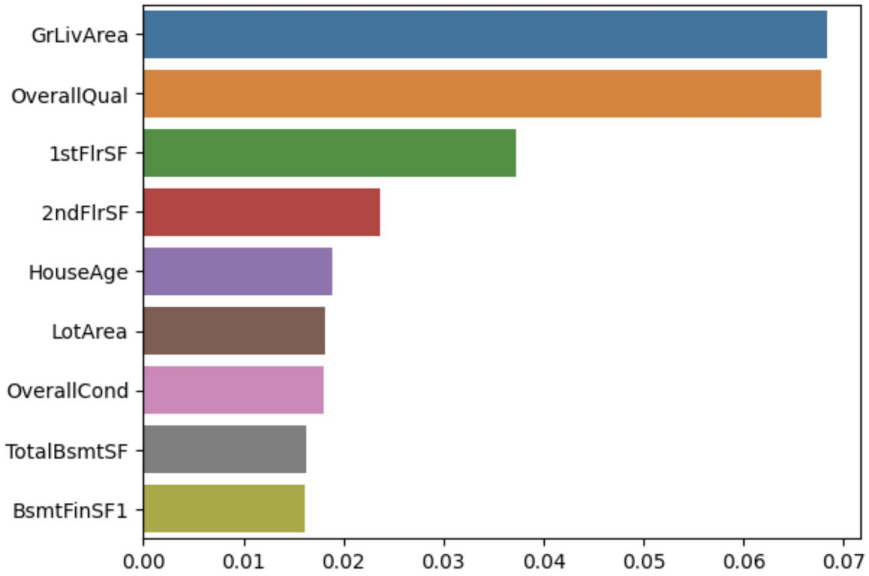
GB:

Trees : 10K
Depth : 3
Sample Leafs : 1
Sample Split : 2
Learning Rate : 0.01

Support Vector Regression: RBF Kernel

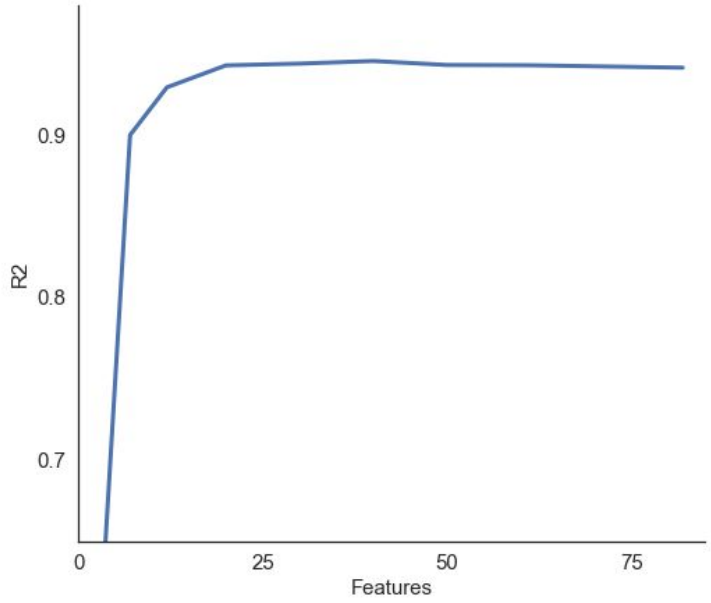


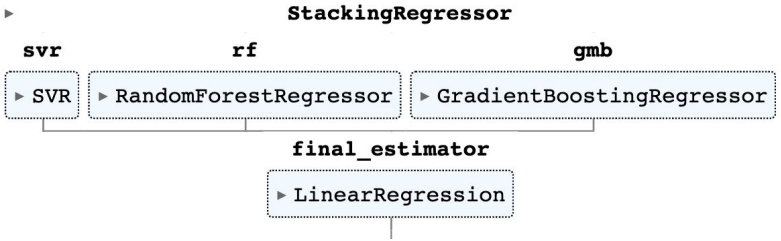
Top Important Features, SVR:



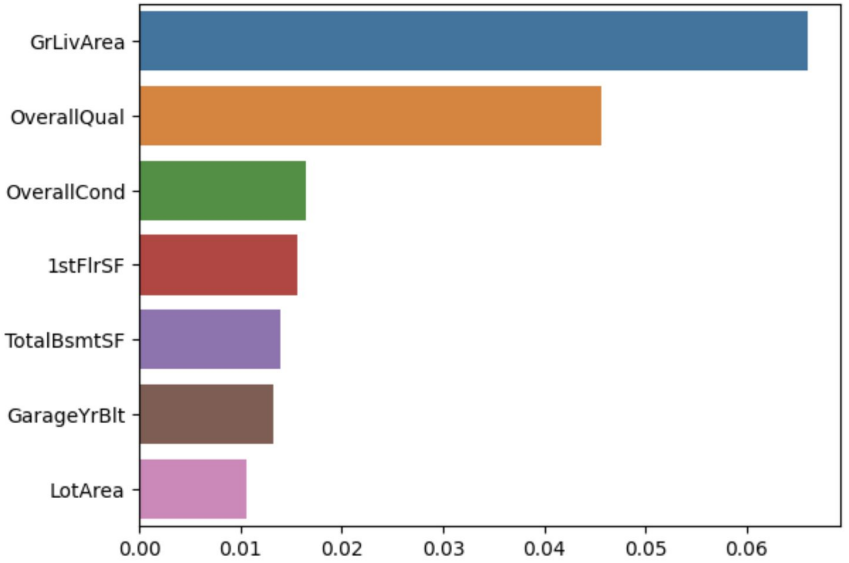
C : 10,
ε : 0.04,
γ : 0.002

Stacked Model: Power Rangers, Unite!





Top Important Features, Stacked:



The **how**.

Closing Thoughts: The most important features from Stacked Model

GrLivArea

Focus on building structure. Choose single-story homes with potential to expand **upwards** to increase square footage.

OverallQual

While looking for **worn-down** homes, pay attention to the building's material quality. Brownie points if they have an exterior with cement bedding.

TotalBsmtSF

While it may be difficult to increase basement square footage, it's **moderately correlated** with basement quality and condition.

But don't take our word for it. See for yourself!

(click to access the playground)

SHINY