How much for the 'Dank Nuggs'?

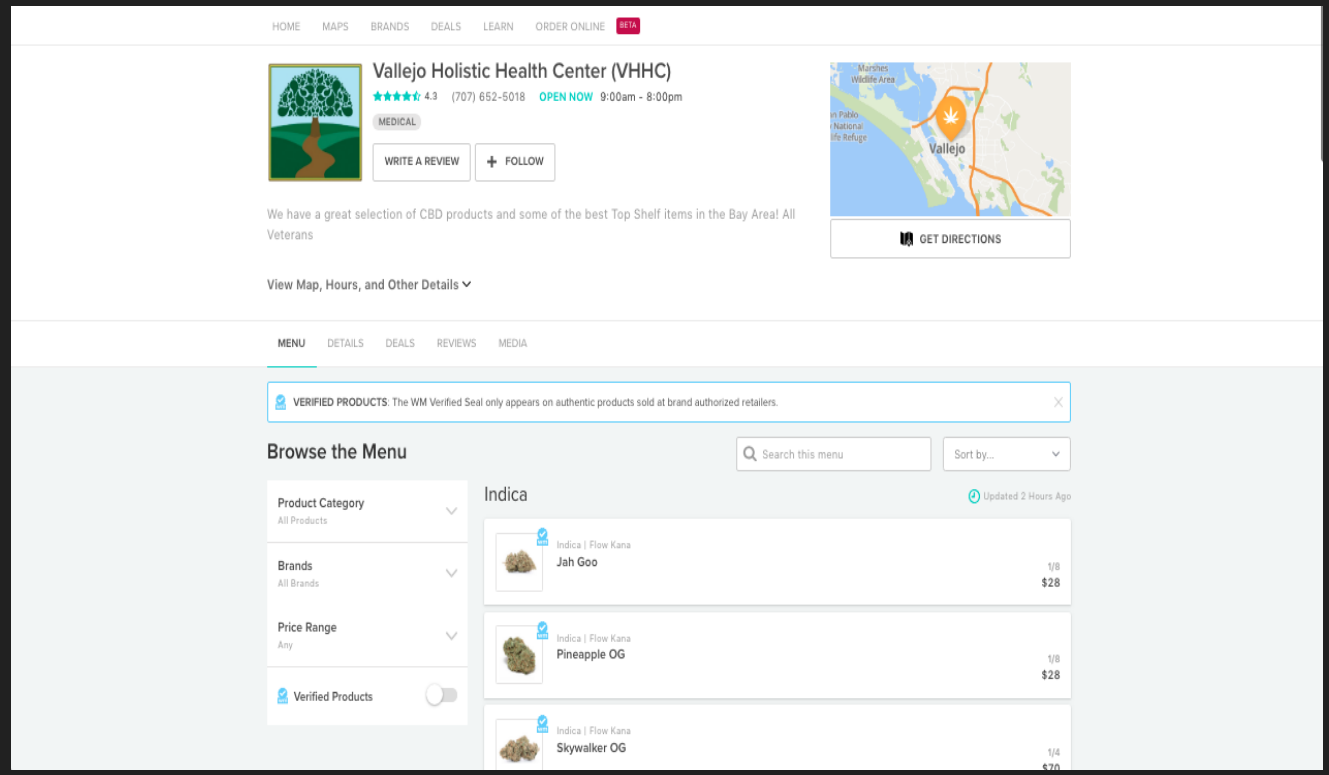
AJ Davis Project Luther

Problem

- Can we predict cannabis flower prices at dispensaries?
- Why relevant?
  - Are dispensaries pricing their flower products optimally?

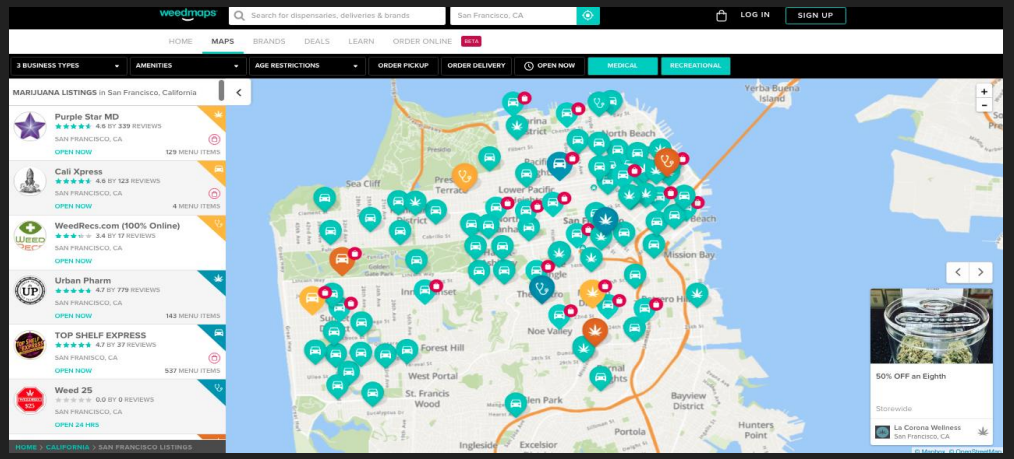
Approach

- Data
- Supervised Machine Learning



Pipeline

- Scrape the data
  - Cities → Geocodes → Dispensary Names → Dispensary Menus
- Identify apriori model specifications
- Check specification scores for various models
- Fit to test data and check for overfitting



Feature Space

- THC Content
- Cross-section of dispensaries
- Dispensary ratings
- Dispensary license type
- Strain name
- Strain type
- Rare Strains

Final Model Specifications

Specification 1

$$Price = THC\% + Rating_D + License_D + \sum_{r=1}^{R} D + \sum_{t=1}^{T} D$$

Specification 2

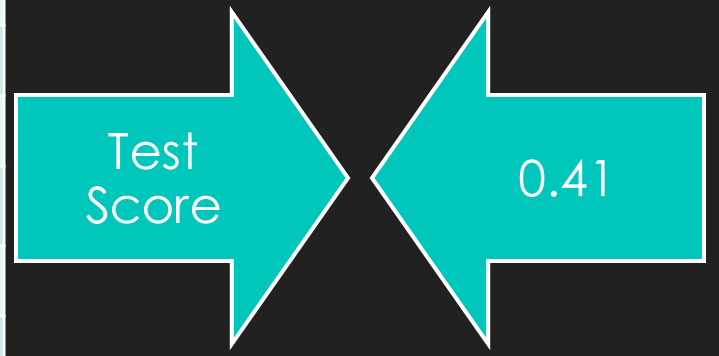
$$\log(Price) = THC\% + Rating_D + License_D + \sum_{r=1}^{R} D + \sum_{t=1}^{T} D$$

Specification 3

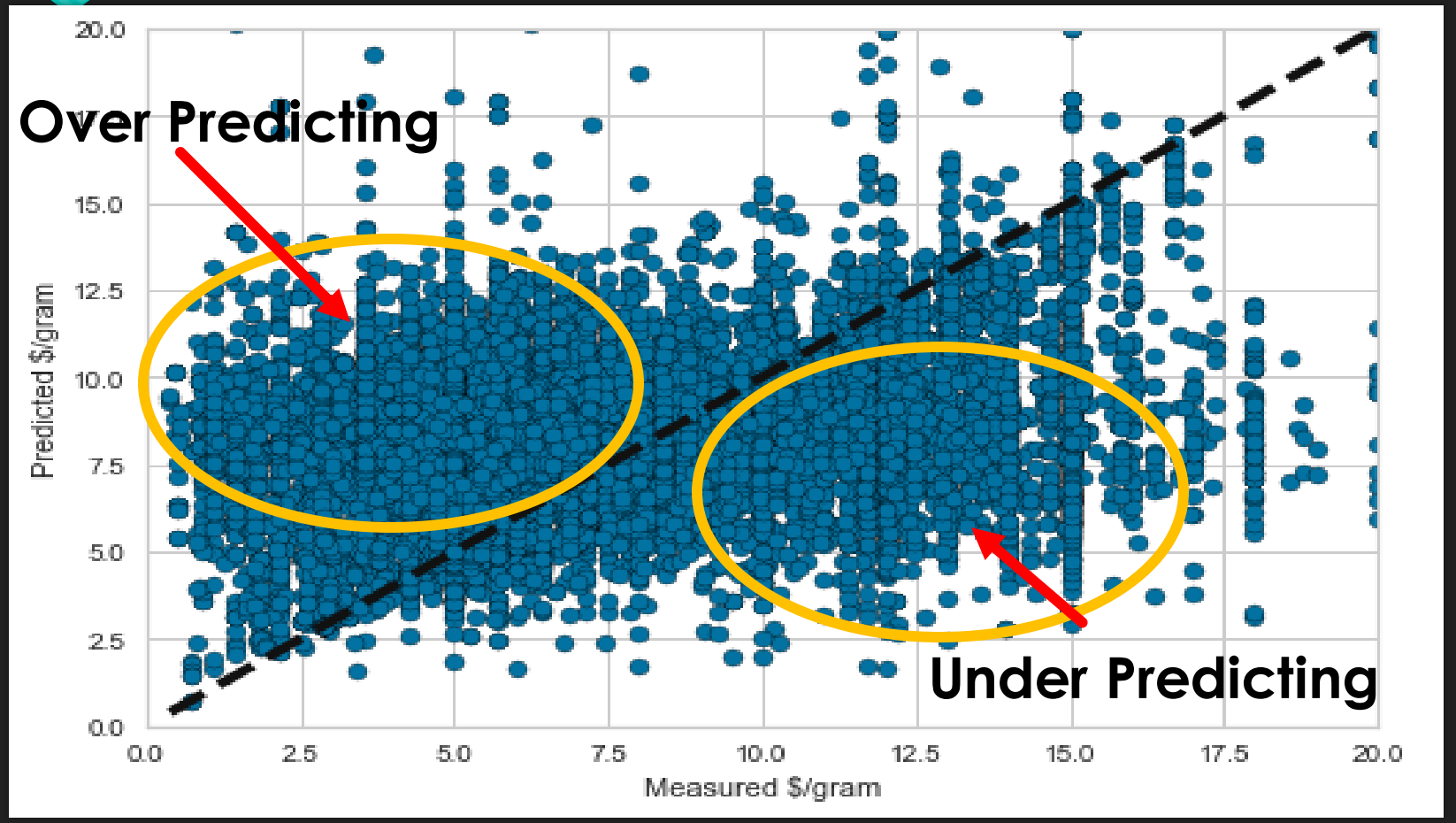
$$Price = THC\% + THC\%^2 + Rating_D + License_D + \sum_{r=1}^{R} D + \sum_{t=1}^{T} D$$

Results

Model (Training)	Specification 1 Score (R2)	Specification 2 Score (R2)	Specification 3 Score (R2)
Linear Regression	0.10	0.13	0.22
Ridge Regression	0.09	-1.44	-8.51
Lasso Regression	0.05	0.00	-0.19
Robust Regression	-0.14	-1.34	-0.86
Epsilon Insensitive Regression	-1.18	-1.42	-2.13
CART	0.30	0.21	0.34
Extra Trees	0.25	0.14	0.21
Random Forest Regression	0.41	0.30	0.48
ADA Boosted Trees Regression	-0.08	0..2	0.16
Gradient Boosted Trees Regression	0.39	0.27	0.40



How well is the best model predicting?



Discussion and Next Steps

- Critical omitted features
  - "Nose"
  - Density
  - CBD?
  - Quantity
  - Demand
- Better string processing for strain data
- Look at other dispensary products
- Hyperparameter tuning