First Price Auctions Bid Optimization with Contextual Multi Armed Bandits

by Jacob Greenbaum, Himani Gadve, Mike Arbuzov

How to win **bidding wars** without overpaying a fortune?

SFGATE

San Francisco home sells for $1 million over asking price

A four-bedroom home in a quiet San Francisco neighborhood shocked real estate agents Richard Woo and Holly Phan when it sold for $1 million...

1 day ago

Money

Housing Market Madness: This 4-Bedroom Home Just Sold for $1 Million Over Asking Price

Whether it's your first home or your next one, Quicken Loan experts can walk you through the process. Click below to consult a mortgage expert...

1 hour ago

CBS San Francisco

'We Didn't Expect A Million Dollars More' – San Francisco Home's Astronomical Overbid Stuns Agents

SAN FRANCISCO (KPIX 5) – While high prices and overbidding have become the norm in the Bay Area real estate market, a home that sold for $1...

3 days ago

Agenda

1. **Tax Rolls to Sales & Feature Engineering**

   @Jacob Greenbaum

2. **Models (QRF, QGBDT, NGBoost) & fine tuning with Optuna**

   @Himani Gadve

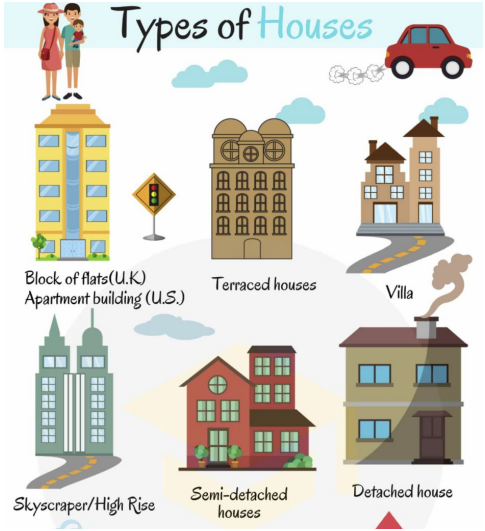
3. **Multi Armed Bandits & Results**

   @Mike Arbuzov

1. Tax Rolls to Sales

How we back engineered dataset and generated features

Defining the scope



Train

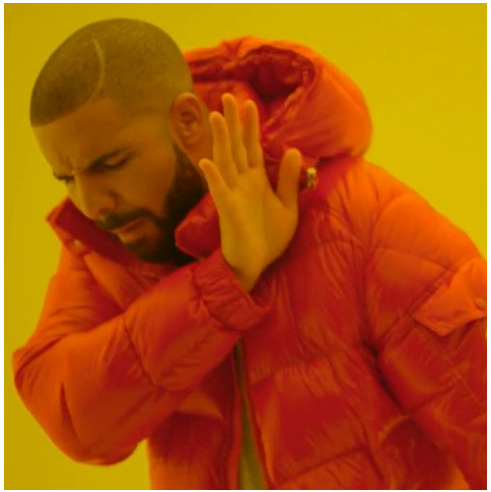
Property Class: Condominium or Dwelling
Use Definition: Single Family Residential
Bathrooms <= 5
Bedrooms < 5
Stories < 5
Property Area(sqft) 500 - 5000
Home Value within 5% - 95% percentile of training

Predict

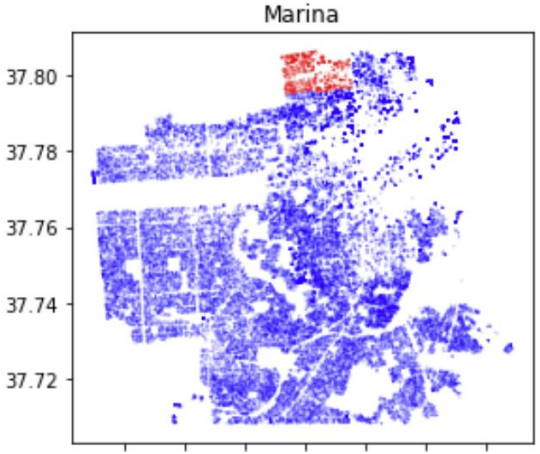
Fixtures
Improvement
Land
Personal Property

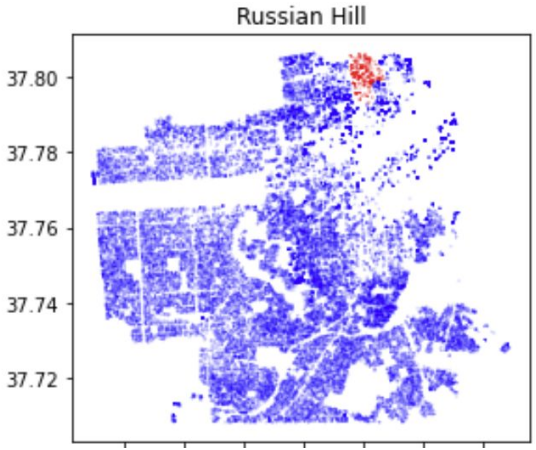
$=$ Assessed Value $\approx$ **Sale Price**

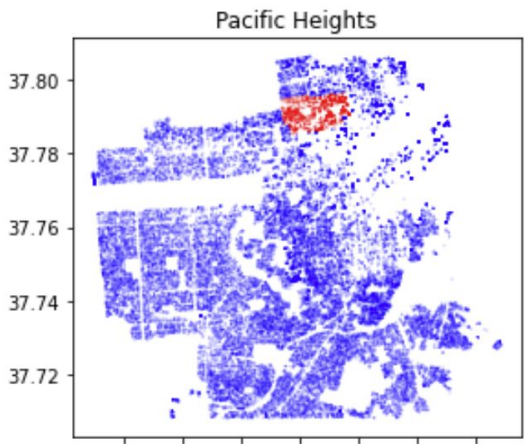
Predict on:
Sale Price?!

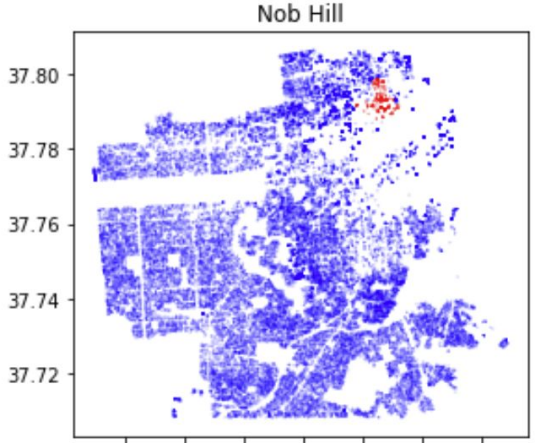


Predict on:
Sqft Price /
Median Sqft Price
in the Neighborhood
for Last 26 weeks



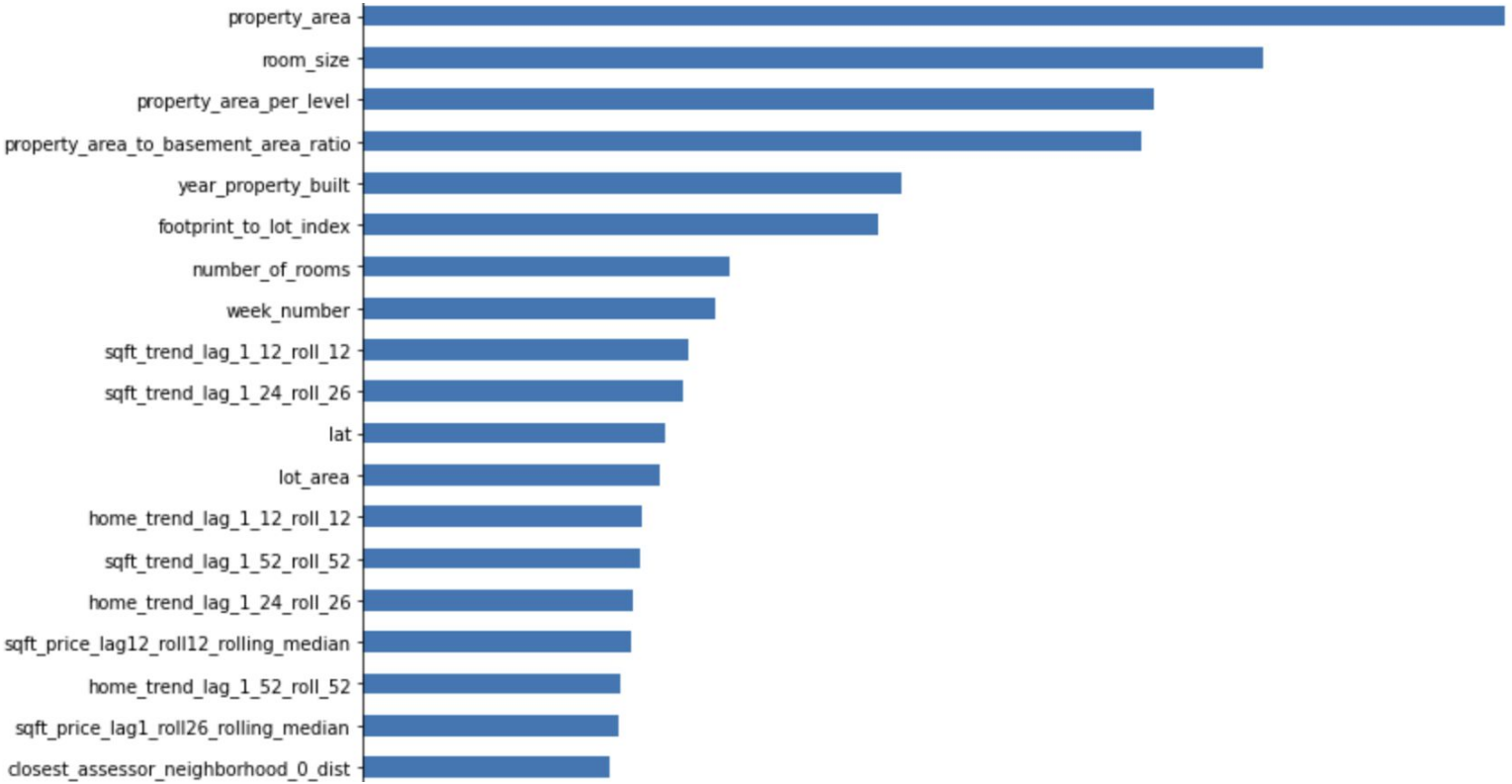






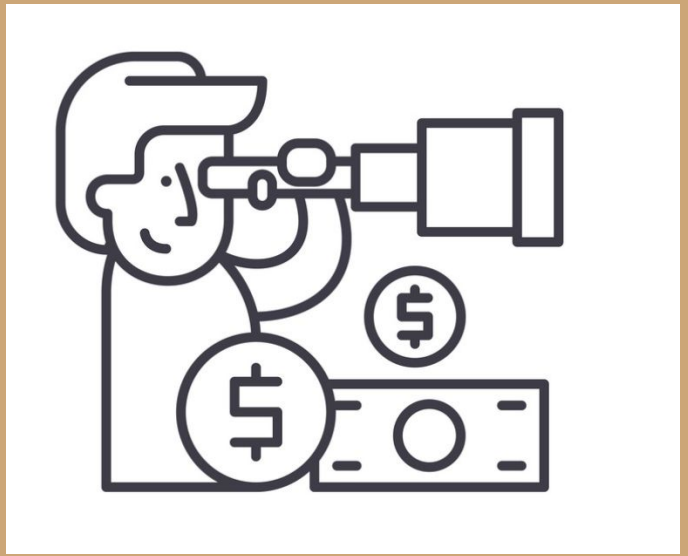
Plots made on: GeoPandas

Valuable features - selected with RandomForest

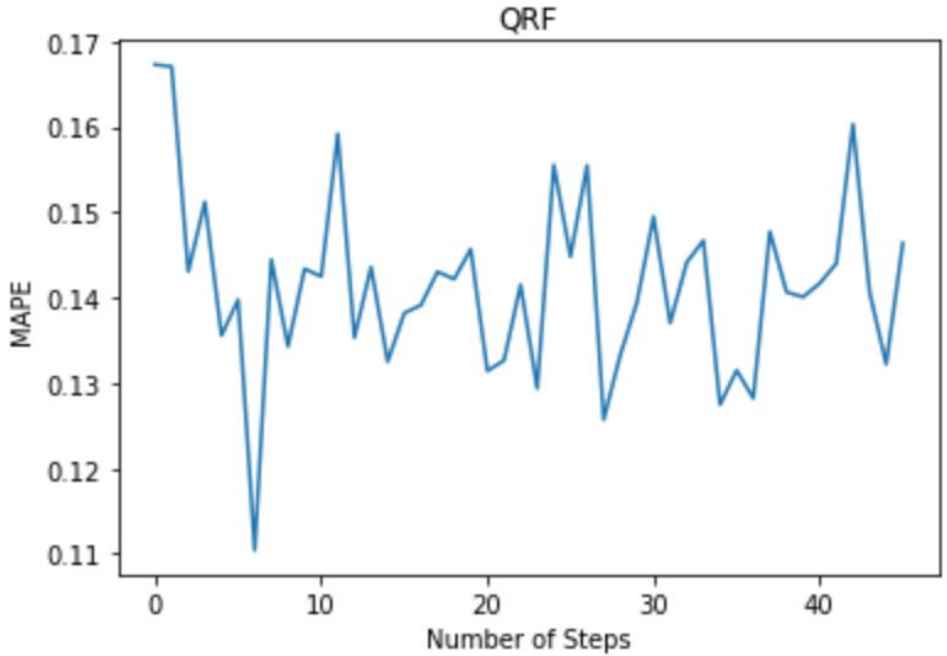


2. Probabilistic Forecasting

Forecasting confidence intervals with QRF, QGBDT & NGBoost



QRF - Random Forest Quantile Regressor: Predicts residual of a linear regressor



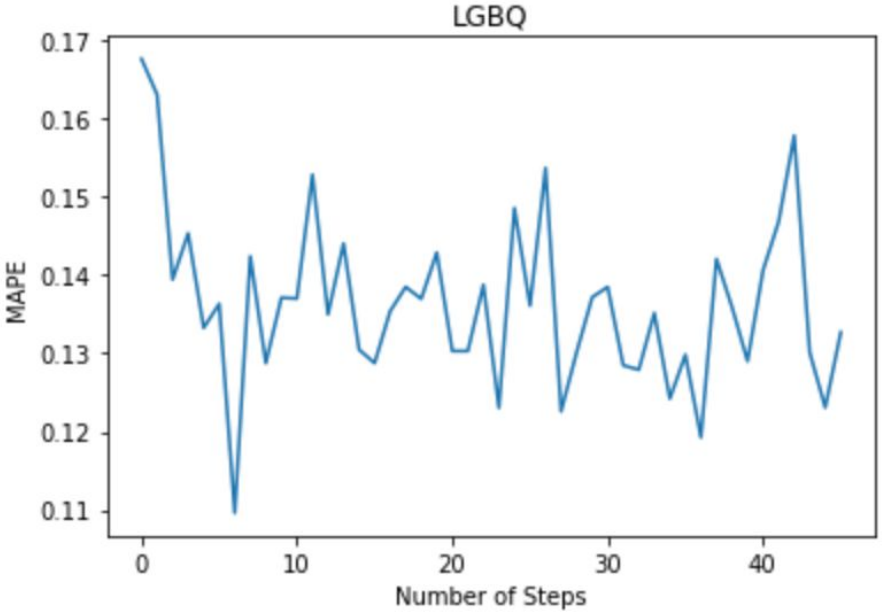
Tuned Hyperparameters:

Number of Estimators: 10

Max_depth: 120

Best MAPE value: 0.11

QGBDT - LightGBM Quantile Regressor: Predict different quantile values



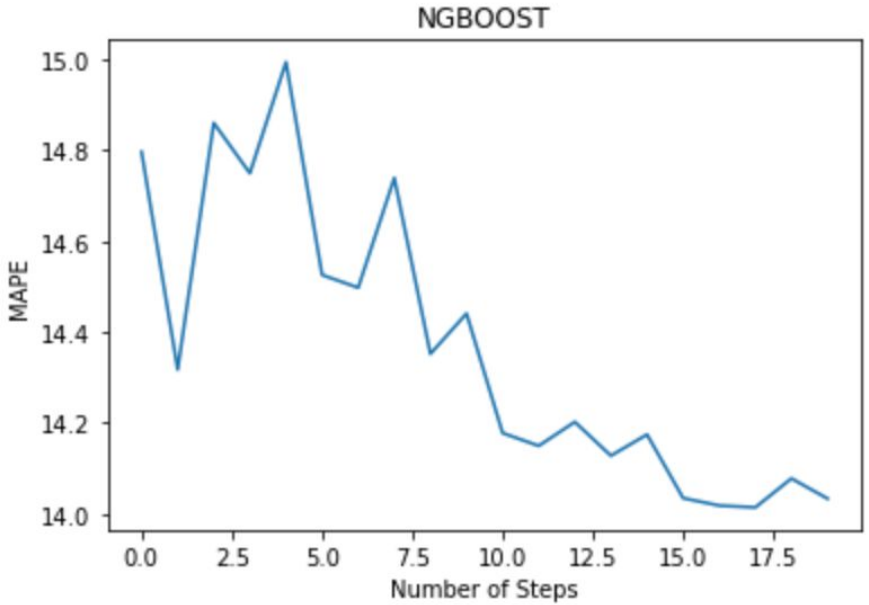
Tuned Hyperparameters:

Max_depth: 4

Learning rate: 0.012

Best MAPE value: 0.11

NGBoost Regressor: Predict mean and standard deviation



Tuned Hyperparameters:

Number of Estimators: 196

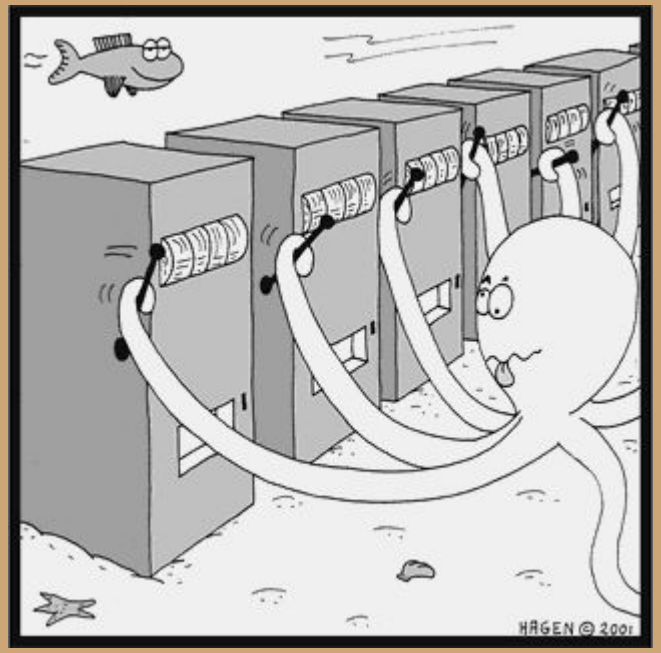
Max_depth: 7

Min_samples_leaf: 25

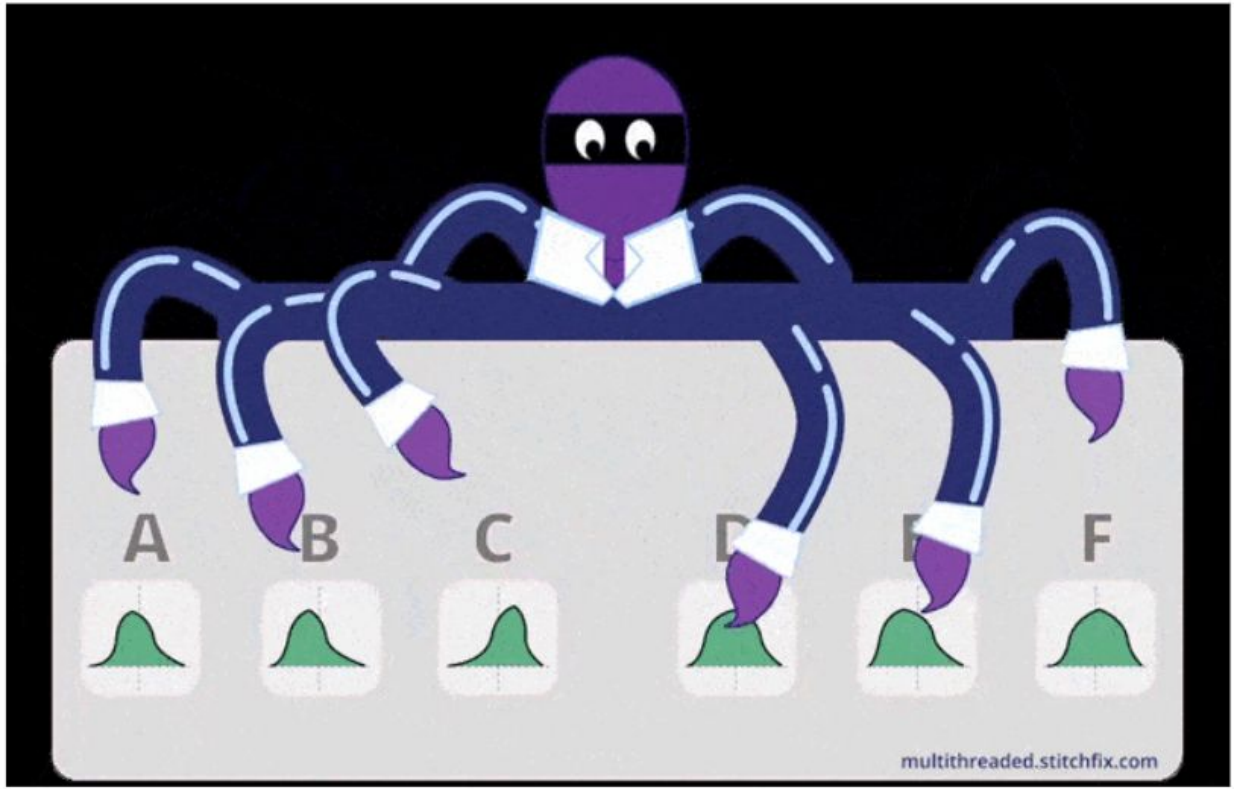
Best MAPE value: 0.14

3. Multi Armed Bandits

Choosing optimal bidding strategy based on previous outcomes



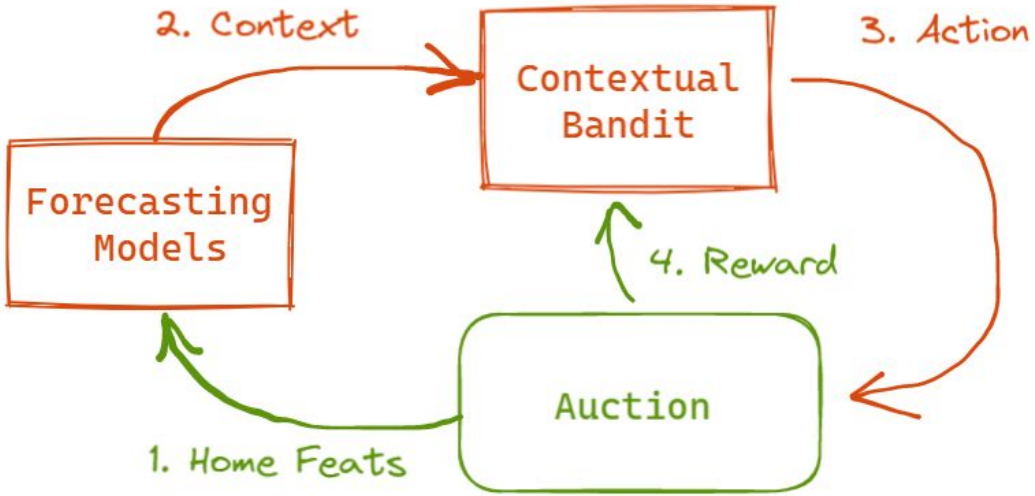
Sampling the rewards



How MABs learn?

Reinforcement Learning:

1. Context generated from Home features
2. Agent receives Context
3. Agent chose random or optimal Action
4. Agent receives Reward
5. Agent adjusts action selection model



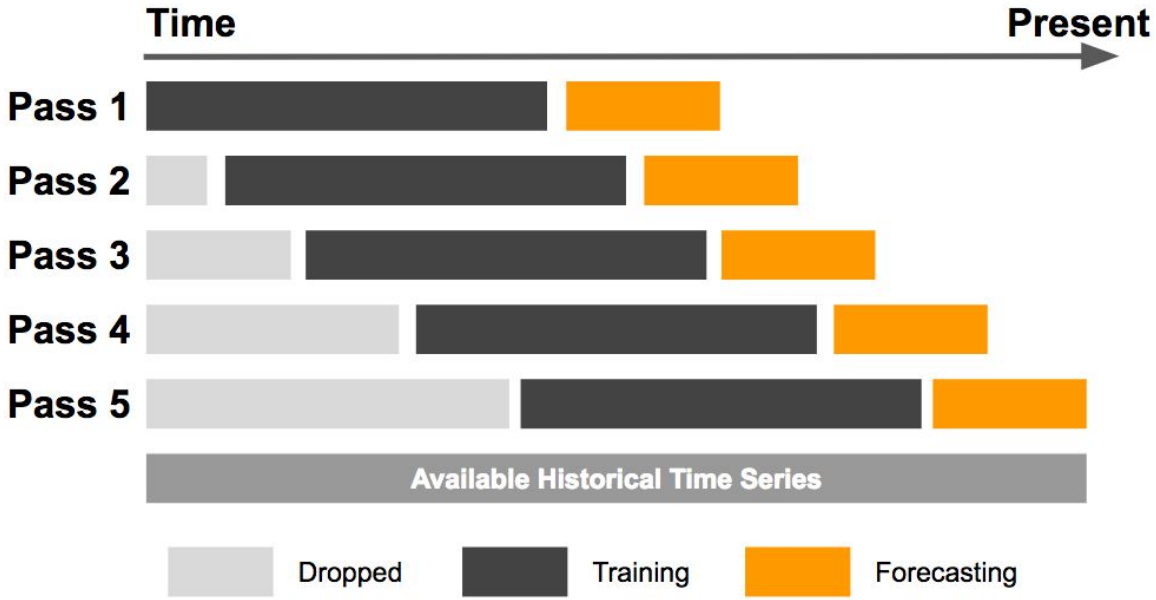
Context

1. Train forecasting models on five years of historical data
2. Create predictions for the homes selling during the following month

{ 'ngb_pred', 'ngb_std',

'QGB_10','QGB_30', 'QGB_50', 'QGB_70', 'QGB_90',

'QRF_10', 'QRF_30', 'QRF_50', 'QRF_70', 'QRF_90'}



Actions

- pick optimal bid adjustment based on the context.

Depending on market situation and forecast precision, MABs strategies range from aggressive bidding to ceasing bidding at all.

```
actions = (
    '-12', '-9', '-6', '-3', '-1',
    '0', '1', '3', '6', '9', '12',
    '-100'
    )
prediction = (
    context['ngb_pred'] +
    context['QGB_50'] +
    context['QRF_50']) / 3
    )
bid = prediction * (100 +
int(action)) / 100
```

Reward = negative Cost

helper artificially reduces error

if no win, lost time loss = 0.5%

elif overbid within thresh%, earn commission loss = - 2.5%

else overbid by more than thresh%, loss = excess %

helper - coefficient to reduce forecasting error; helps examine how MABs strategies depend on forecast quality

```
threshold = 0.05
diff = ((bid - target) / target) * helper


if diff < 0:
    return 0.005
elif diff < threshold:
    return -0.025
else:
    return (diff - threshold)
```

Results - homes purchased / avg overbid %

MAPE	0%	1.5%	3%	6%	10%	14%
MABs	16k / 1%	16k / 3%	1k / 5%	0.8k / 9%	0.8k / 13%	0.7k / 18%
Median	8k / 0%	8k / 2%	8k / 3%	8k / 7%	8k / 12%	8k / 17%
- 5%	0k / na	0.4k / 1%	2k / 3%	4k / 6%	5k / 11%	6k / 16%
+5%	17k / 5%	17k / 5%	16k / 6%	13k / 8%	11 k / 13%	10 k / 18%

Zestimate median error is 3.5% in San Francisco, CA
(Based on 10,000 homes with Zestimates)

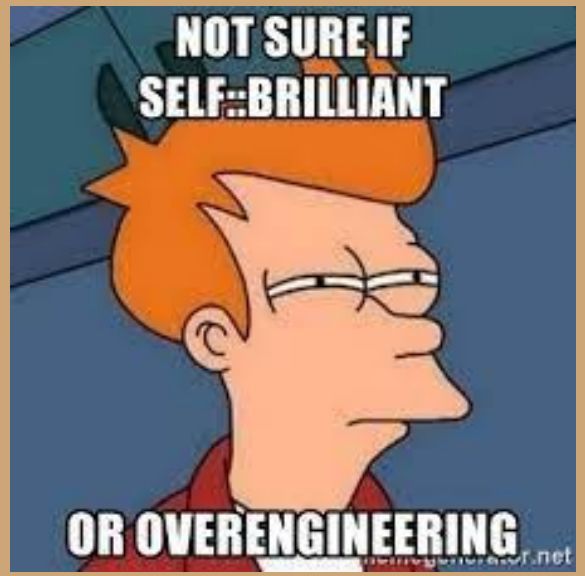
MAB Strategies

Prediction Error	Strategy	Purchases / Bids
0.0%	+ 1% (min premium)	16 / 17
1.5%	+ 3 % (premium > avg error)	16 / 17
3.0%	- 9% (lowball)	1 / 17

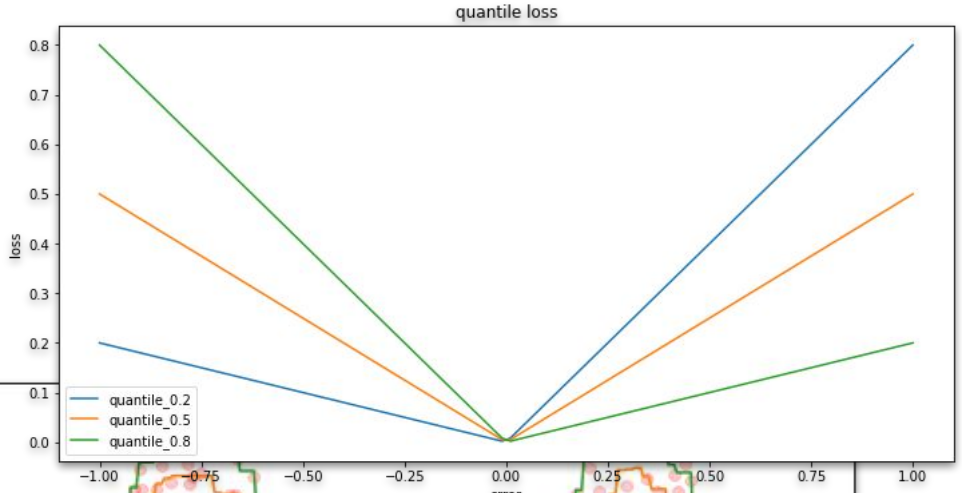
If **MAP > threshold**, Multi Armed Bandits figure out to **refuse to bid**.

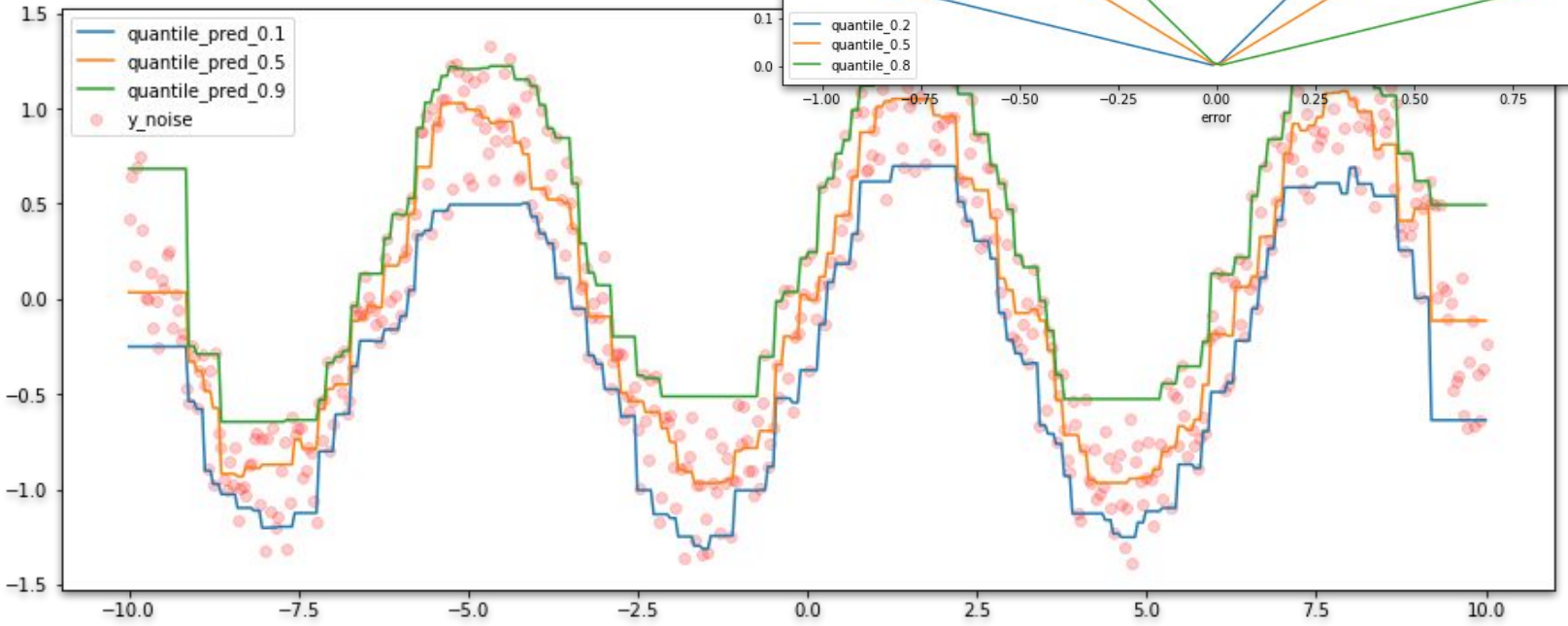
4. Appendix

Quantile Regression
& DASK Feature Engineering



Quantile Regression - estimates prediction intervals using quantile loss





DASK - Parallel Feature Engineering

140 features has been created with DASK for parallel computations:

- Geospatial - distance to neighborhood centers, closest neighborhoods
- Time Series - lags, rolling medians, and trends
- Property - footprint to lot ratio, area per level, lot to property ratio and others