



PRICE PREDICTION DATATHON FME 2024

Alba Ungueti, Martina Massana, Raquel Jolis and Maria Sans



Challenge statement

The task is to create a **price prediction model** known as an **Automated Valuation Model (AVM)**.

The goal is to assist real estate agents in **estimating the best price for buying or selling a property based on historical data**. This recurring challenge involves developing a model that uses the characteristics of a property to provide an accurate market value estimate.



Table of contents

01

Pre-processing & Data
cleaning

02

Model trial

03

Final result

04

Conclusions



01

Pre-processing & Data cleaning



Data augmentation

Added **data of Median HouseHold** Income from:

<http://www.usa.com/rank/illinois-state--median-household-income--zip-code-rank.html>

Merged by Zip Code to the original dataset



Median Household Income	Zip
179922	60043
172539	60022
166667	60029
157212	60521
155750	60604
...	...
16706	62914
16500	62204
14513	62701
14432	62059
8993	62523



Preprocessing

- **Binarization:** Parking
- **Categorical one-hot encoding** using **kNN**: Features reso
- **Dimensionality reduction** based on **correlation with ClosePrice**: Features reso, coordinates
- Numerical **imputation** using the **mean**: c1c6 summary Property
- **Numerical imputation** using **kNN**: Median Household Income



02

Model Trial



Models

1. Linear Regression
2. Lasso Regression
3. Ridge Regression
4. K-Nearest Neighbors
5. Random Tree
6. Random Forest
7. Random Forest Optimization
8. Extra Trees



03

Final Results



Models

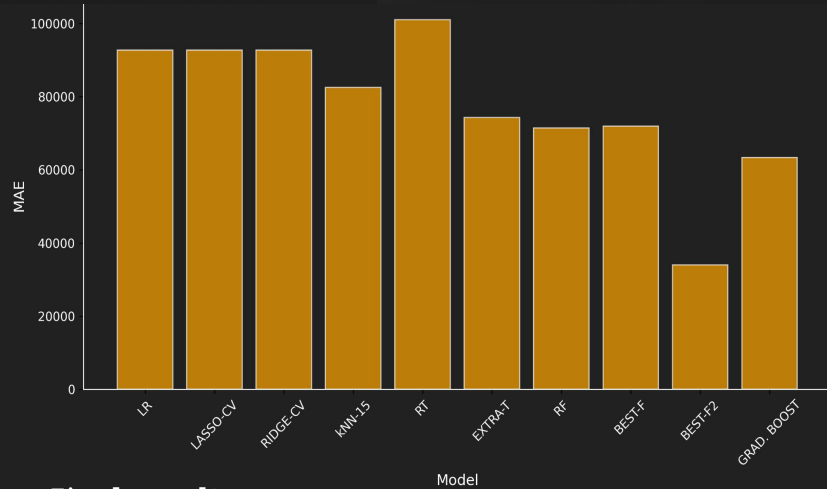
1. Linear Regression
2. Lasso Regression
3. Ridge Regression
4. K-Nearest Neighbors
5. Random Tree
6. Random Forest
7. Random Forest Optimization
8. Extra Trees
9. Embedding Gradient Boosting

MODELS	R2	MSE (10 ⁹)	MAE (in thousands)
LRX	0.65	18.31	92,73
LASSO-CV	0.65	18.31	92,77
RIDGE-CV	0.65	18.31	92,77
kNN-15	0.70	15.84	82,56
RT	0.54	24.30	101,06
EXTRA-T	0.76	12.71	74,27
RF	0.77	11.96	71,43
BEST-F	0.77	11.94	71,10
GRAD. BOOST.	0.83	9.13	63,36



Models

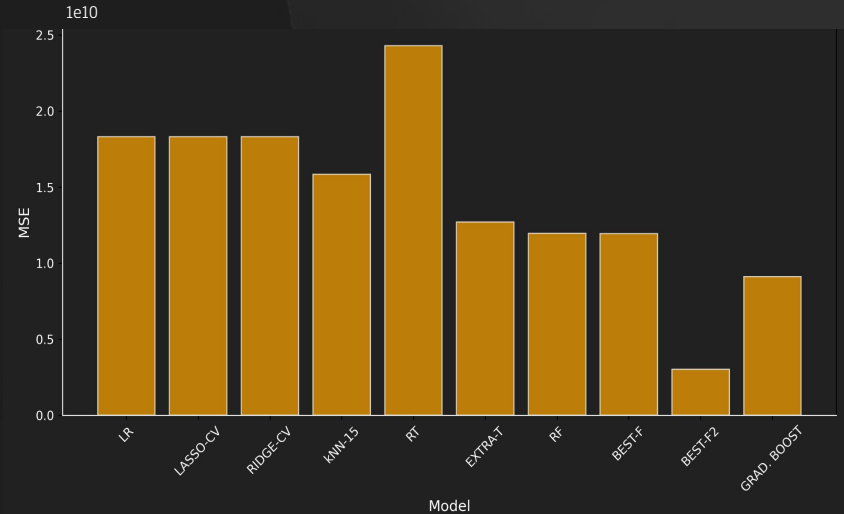
Comparison of MAE per model



Final results

MAE: \$98,141

Comparison of MSE per model



MSE: \$66,002,187,355



04

Conclusions

