Exploring real estate prices in New York City

Wiktor Jakubowski

Mikołaj Gałkowski

Hubert Bujakowski
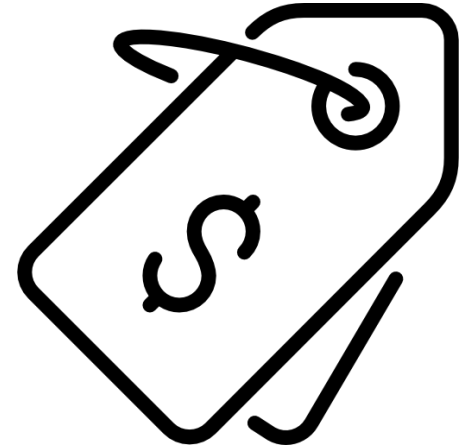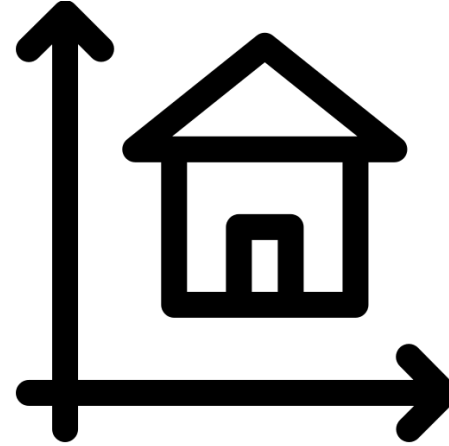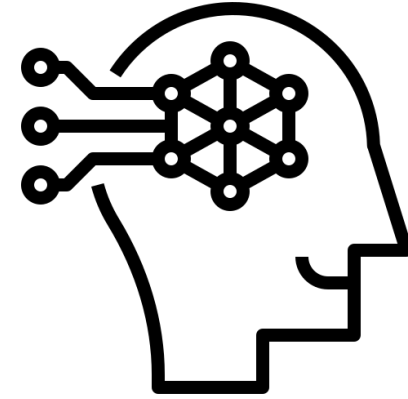
Łukasz Tomaszewski

Problem

Rapid growth of average prices in Metropolis like New York City
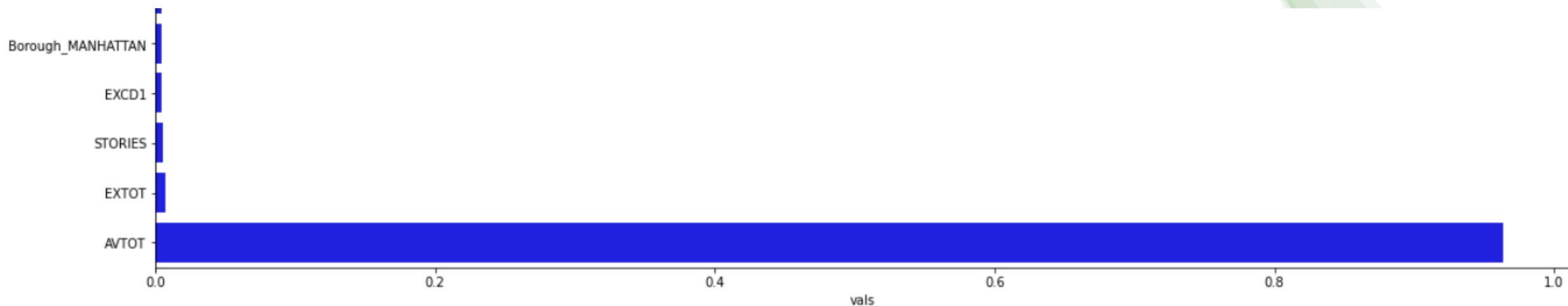
Dataset

- Regression model task

- Dataset about NYC real estates

- Target variable: price

Model

- Our main goal was to create the most interpretable model

- Chosen algorithms: random forest & xgboost
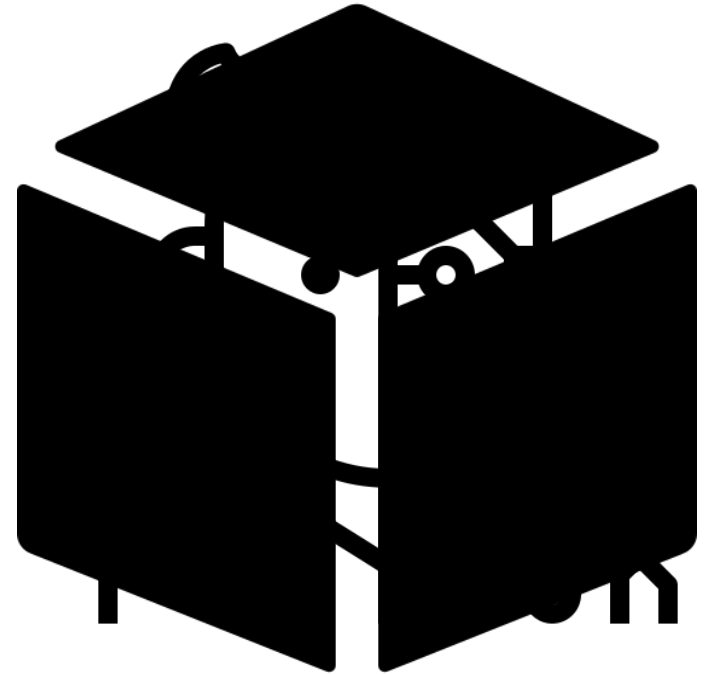
- Chosen metrics: MSE & MAE

Feature Engineering

LEAST IMPORTANT	MOST IMPORTANT
Building class	Actual Total Value
Tax class	Number of storeys
Year of construction	Borough

How did we use it

- Not a BlackBox

- Easy to understand for humans

- Not overtrained

Final results

- Best model before HPO MAE was around 120k $

- Final model MAE was 100k $ (8% of avg property price)

Concept

- Brainstorm – helping customers / sellers

- Trying to increase peoples awareness
of possible exploits

Application - Main Idea

- Visualize the map of NYC
- Show the possible real estates to buy
- Compare the actual price scrapped from the Internet and the prices evaluated by our model