Hempriser

Scraping 1.3 million home prices

and getting away with it

Pierre Le Fevre, Emil Karlsson

PyData Stockholm, 2024-02-22

Who?

- Emil Karlsson, Pierre Le Fevre
  - KTH, MSc distributed systems

- ID2223 Scalable ML

Why?

- Fun technical challenge
- Easy to compare to existing predictors

booli!

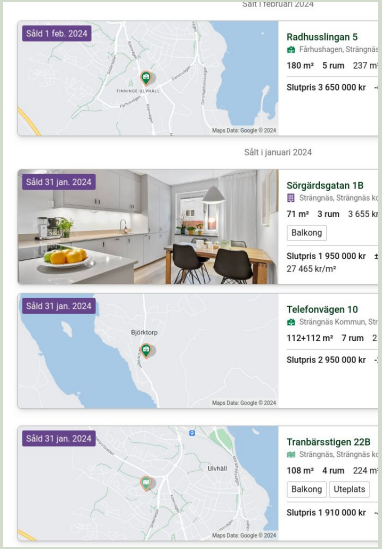
Blok

Backfilling

Locations

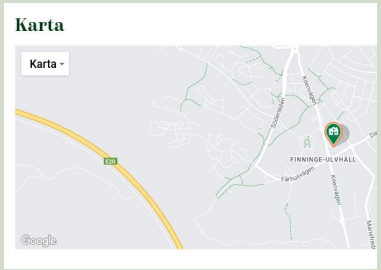
street
stock
gothe
…

Property listing

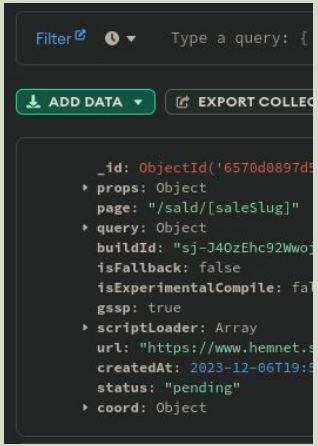


Metadata

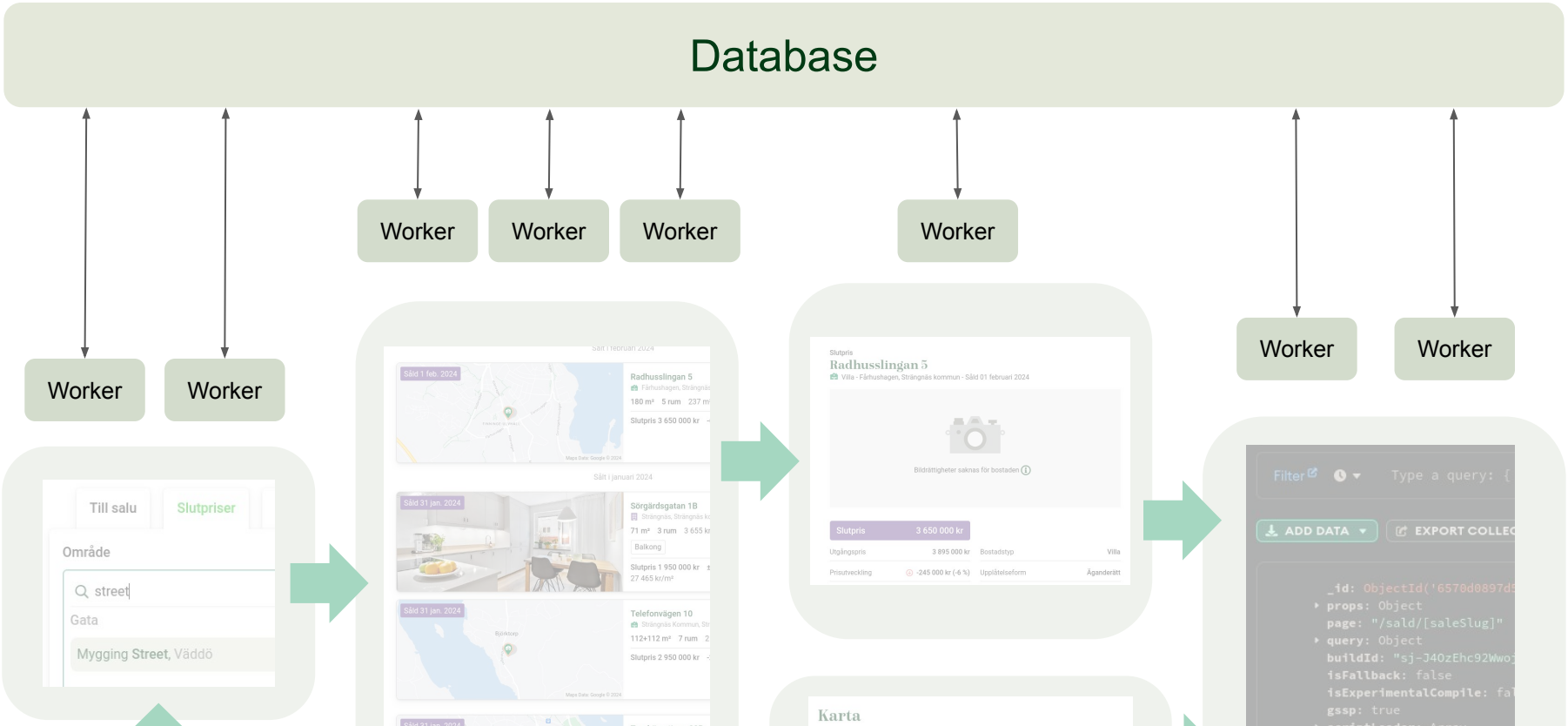
Coordinates



Raw data



Parallelization



Model training

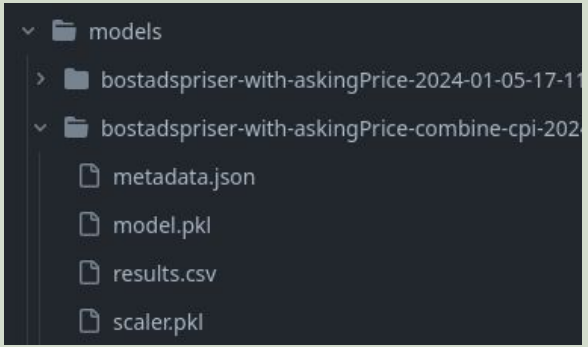
Setups

without-asking-price
Features: cpi, fee, age ...

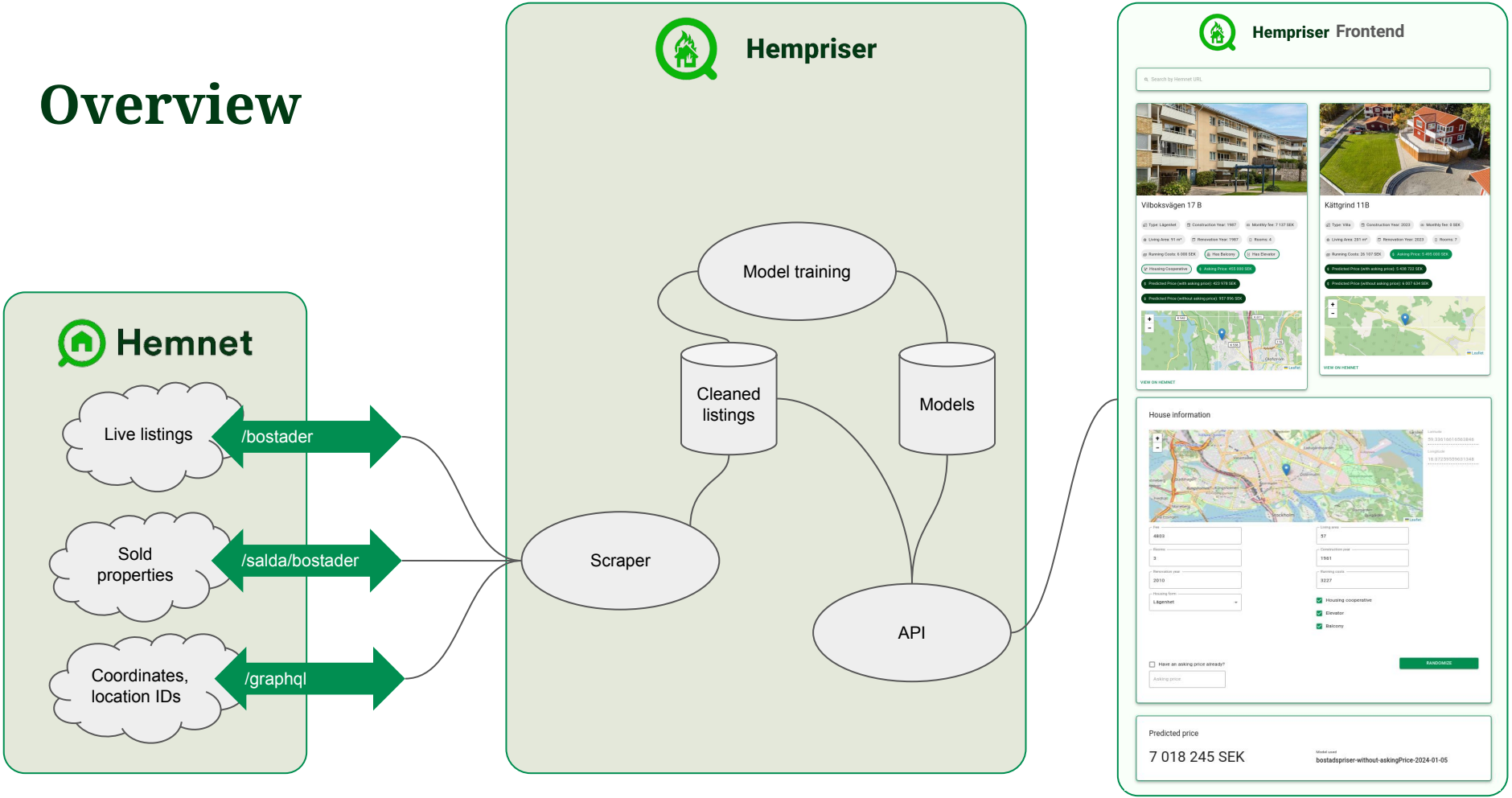
with-asking-price
Features: askingPrice, fee ...

Regressors
Linear
Gradient Boosting
n_estimators, max_depth ...

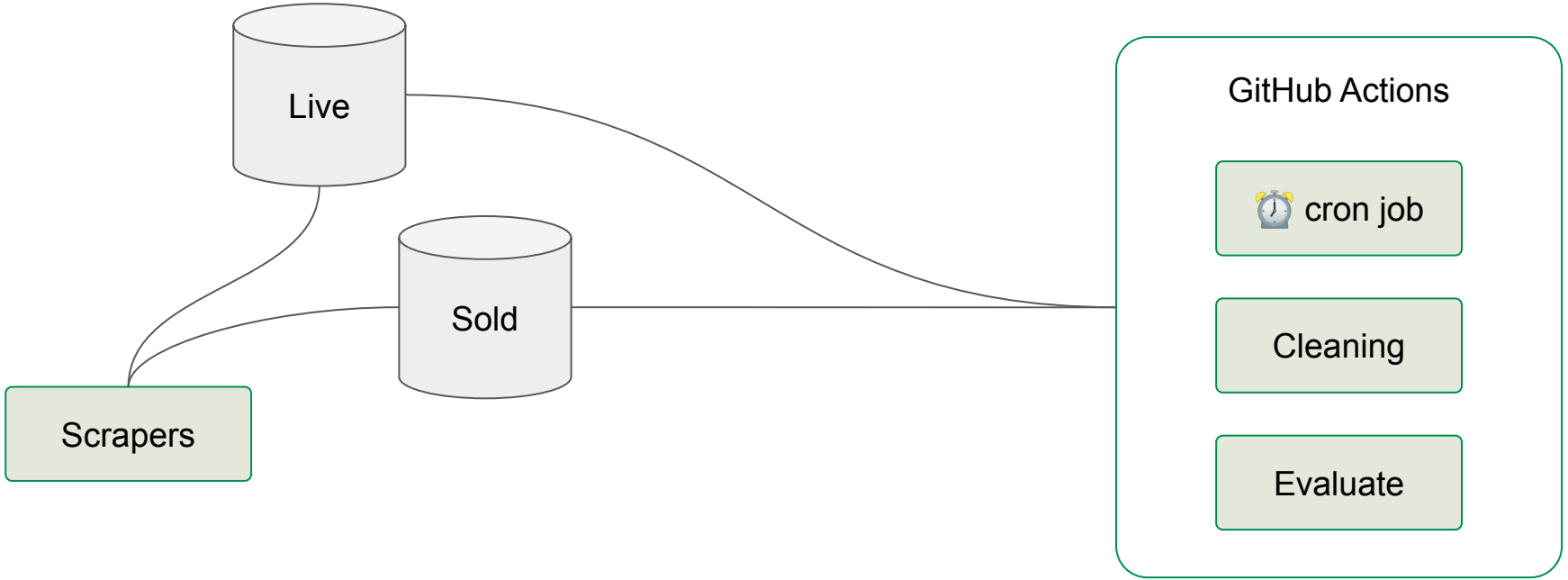
scaling
transforming
HPT & training



Overview


Hempriser
Hemnet
Live listings
/bostader
Sold properties
/salda/bostader
Coordinates, location IDs
/graphql
Model training
Cleaned listings
Models
Scraper
API
Hempriser Frontend
Vilboksvägen 17 B
Kättgrind 11B
House information
Predicted price
7 018 245 SEK
bostadspriser-without-askingPrice-2024-01-05

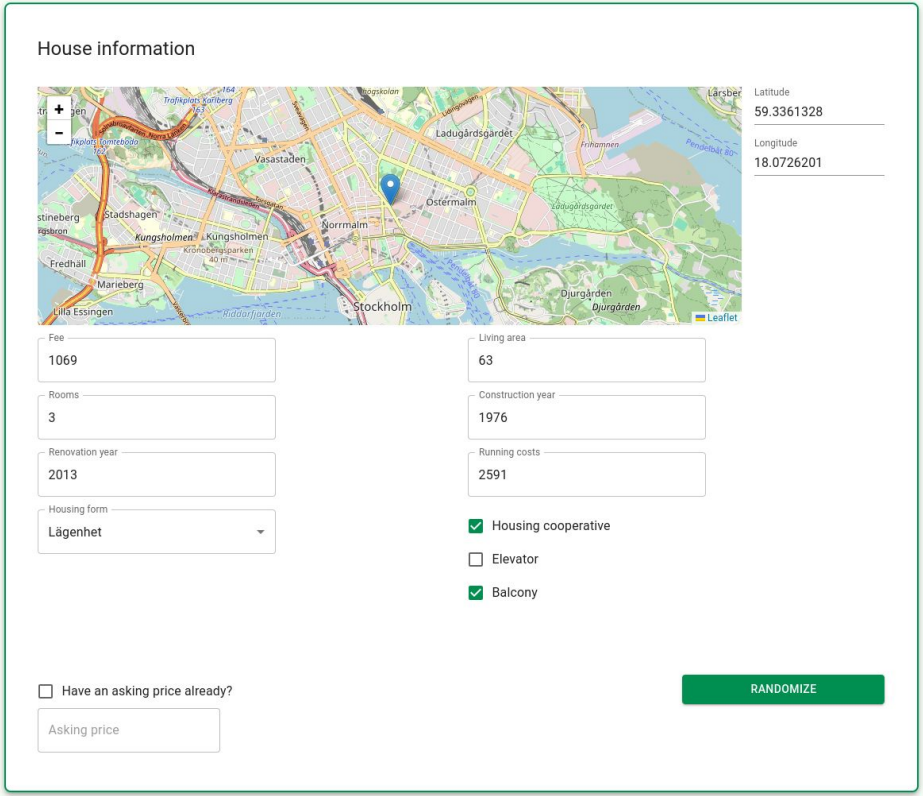
MLOps pipeline



Inference

Different models based on features

Small model - fast response



Live demo

Nice price

Open source

Check it out for yourself!

Code (and dataset 👀) →

github.com/pierrelefevre/hempriser

Thanks!