263-5052-00L FS2024
Interactive Machine Learning: Visualization & Explainability

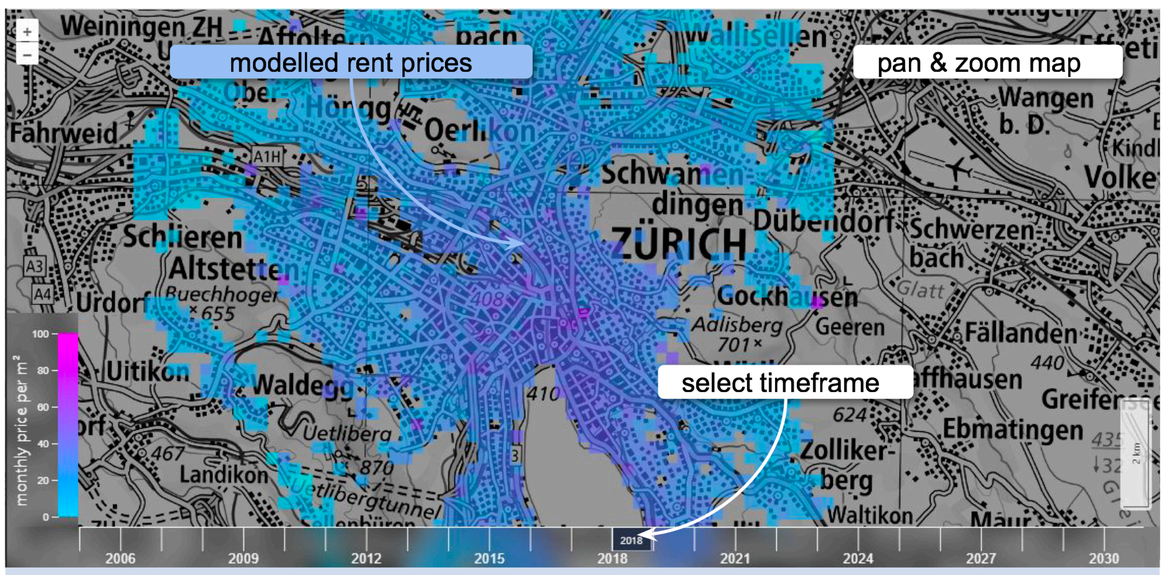
GentrificationSim

Interactive exploration of housing rental prices in Zurich

Aleksandra Apolinarska[1], Yanick Bachmann[1], Xiyi Chen[1], Florian Schweingruber[1]
[1]D-INFK, ETH Zurich

http://c2-zurich-gentrification-exploration.course-xai-iml24.isginf.ch/zurich/2024

Introduction



Our project seeks to uncover and visualize pricing dynamics and gentrification developments in Zurich's housing rental market, along with predicting future trends, emphasizing uncertainty. It is aimed for lay users and general public and strives to make complex data accessible and actionable. We trained a gaussian process model that generalizes past price values to forecast future trends. For explainability, we leverage its ability to model uncertainty, and overlay predictions with real data points.

Data

- Dataset of ca. 325'000 records of online rental ads for houses and apartments in Zurich from 2005 until 2024 (courtesy of Noveletta DataHub [1])
- Available features: geo coordinates, ad date, price/month/m2, floor area, number of rooms, year of construction
- Open-source maps from Swiss Geoportal as a base layer for the visualization in the user interface.

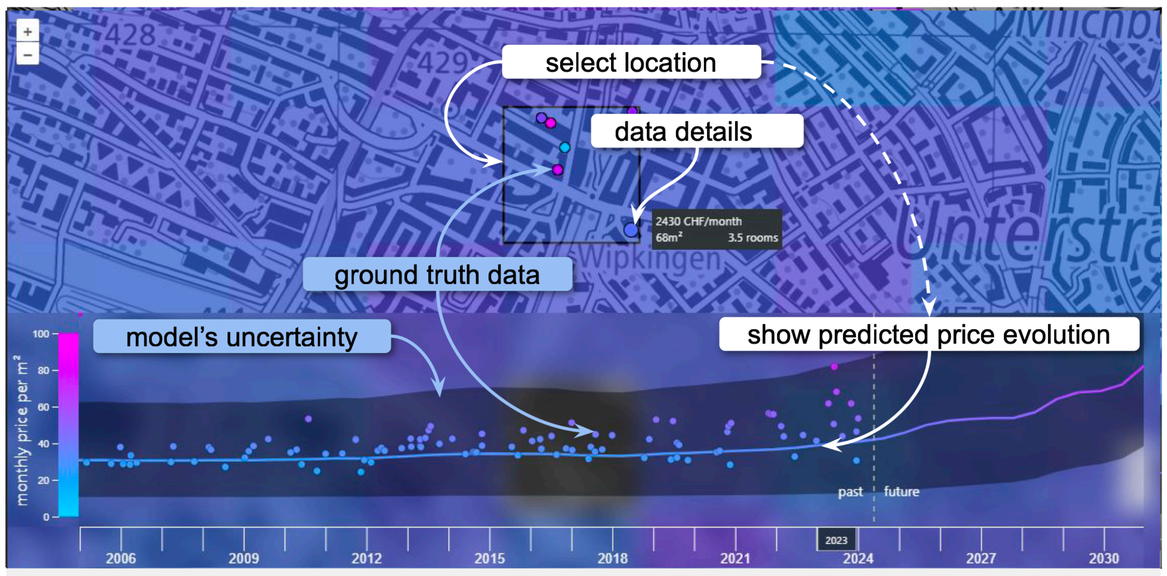
VNNGP model

- We train a Variational Nearest Neighbour Gaussian Process (VNNGP) [2] model to obtain spatio-temporal (based on location and data) prediction of rent prices.
- The model can also express uncertainty about the predictions and make predictions for timepoints in the future.

Implementation

- Frontend Tech Stack: Svelte, OpenLayers, d3.js
- Backend Endpoints: get data sample from dataset, get price prediction from model, get timeseries for a given location
- Latency Optimization: for seamless interaction and to offload computation, precomputed price predictions are stored in backend, tile aggregation is done in backend upon initialisation

User Interface



Tasks and Navigation:

- The UI is a single-page website with an interactive map of Zurich.
- We overlay an animated, time-dependent color map (currently selected year) of rental prices obtained from the model, from 2005 until now and into the future until 2030.
- The user can select a tile on the map to reveal evolution of the rent price within this area over time, at the bottom of the page, predicted by the model and showing the confidence interval.
- The line plot is overlaid with a scatter plot showing data points in the neighbourhood sampled from the dataset.
- The same data points are also displayed on the map. The user can hover over a data point within the highlighted tile to reveal details about this data record in a tooltip.

Visual principles:

- "overview first, details on demand" principle
- showing multidimensional data in different, interconnected views

Explainability

- Visualizing uncertainty of the model predictions. High variance may also signify increase pricing dynamics and/or model's.
- Showing prices only for areas where data was available in the color map, avoiding predictions over irrelevant areas (e.g. no housing or no rental market).
- Showing records sampled from the database in overlay with the prediction. This reveals the density of the data points for given location and timestep, and helps the user adjust their expectation about the accuracy of the prediction.

References

1. Noveletta DataHub https://www.linkedin.com/company/noveletta/
2. L. Wu, G. Pleiss, and J. Cunningham. Variational nearest neighbor gaussian process, 2022. doi: 10.48550/arXiv.2202.01694 2