Italian House Price Prediction (Houseplus)

JACOPO MARAGNA ABDELRAHMAN SALEH
jmaragna | amsaleh @kth.se

December 6, 2024

1 Problem description

This project aims to create a machine learning system for predicting house prices in Italy. It will feature a user-friendly interface where users can input property details, such as square footage, floors, and location, to receive accurate price estimates.

The predictive model will be trained on historical data from Idealista and insights from the Osservatorio del Mercato Immobiliare (OMI), Italy's Real Estate Market Observatory. Predictions will combine user inputs with additional features from OMI to improve accuracy.

To ensure the model remains relevant and reflects market trends, it will be updated weekly with newly scraped data. This continuous update process will enable the system to capture subtle fluctuations in Italy's housing market.

The project is made possible by Houseplus, an Italian startup owned by the 86k holding company. Houseplus will provide exclusive access to high-quality, processed data sourced from Idealista, ensuring a robust foundation for model training and prediction.

2 Tools

- Hopsworks: feature store and model registry.
- Modal/GitHub Actions: Weekly data fetch and model retraining scripts.
- HuggingFace: real-time inference and GUI server hosting.
- Gradio: GUI creation.

3 Data

- Idealista: High-quality, processed data sourced from Idealista, made retrievable through a MongoDB connection.
- OMI: Historical data imported into the system as external files, updated every six months. The model will leverage data collected since 2004 for training.

4 Methodology and algorithm

The approach involves training multiple regression models and comparing their performance. Hyperparameter tuning will be incorporated to optimize each model. Ultimately, the best-performing model will be selected, or an ensemble of models may be used to achieve the most accurate predictions.