2026

Artificial Intelligence Based Automated Real Estate Valuation

Monika Petrulevič, prof. dr. Simona Ramanauskaitė

Vilnius Gediminas Technical University, Department of Information Technologies

30 682
records in the combined dataset of transactions and advertisements

4
Compared AI models: KNN-R, Ridge, SVR, and MLP

0,715
best coefficient of determination R^2 (SVR)

382,84 €/m²
lowest mean absolute error MAE (SVR)

–83 %
price allocation error compared to allocation based on area

01 Introduction

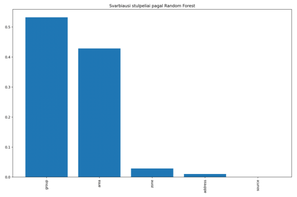
Traditional real estate valuation methods (the comparative, cost, and income approaches) are widely used; however, they are **subjective, time-consuming, and process a limited amount of data**. Furthermore, a single transaction often involves multiple properties, with the individual prices for each remaining unknown.

AIM
Increase the accuracy of automated real estate valuation by combining data from transactions and real estate listings, and applying AI models.

02 Data and methods

1. **Data collection**
13,786 transactions (since 2018) + 23,477 real estate listings (web scraping)
2. **Price allocation**
The transaction amount for multiple items is allocated based on historical proportions
3. **Value zones**
address → coordinates (Google Places API) → zone (REGIA)
4. **Preprocessing**
New features, IQR clipping, RobustScaler, log(1+y)
5. **Model training**
KNN-R, Ridge, SVR, MLP · GridSearchCV + TimeSeriesSplit
6. **Evaluation**
Latest 20% of transactions · MAE, RMSE, R^2, computation time

03 Key features



Random Forest analysis on a normalized dataset: the **property group, area**, and **value zone** are the primary determinants; therefore, the zone was also determined automatically for the listings.

04 Model result comparison

SVR achieved the best overall result on the test set and explains approximately **71.5% of the price variation**. The linear Ridge model was the least accurate.

Table 1. Final model results on the test set (most recent transactions).

Model	MAE (€/m²)	RMSE (€/m²)	R^2	Training (s)	Prediction (ms)
SVR	**382,84**	**706,19**	**0,715**	**67,02**	**0,32**
MLP	406,88	744,23	0,683	15,20	0,003
KNN-R	422,21	775,14	0,656	0,046	0,088
„Ridge“	470,90	835,46	0,601	0,066	0,003

Accuracy depends not only on the algorithm but also on **the composition of the dataset**:

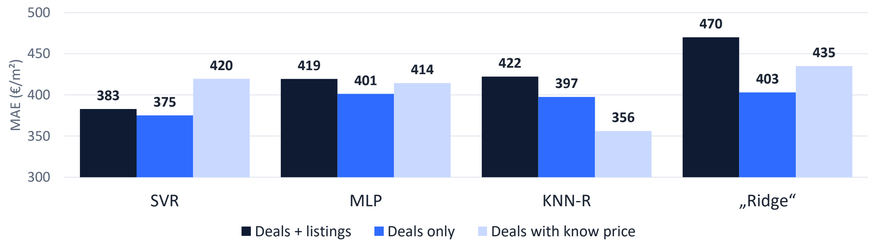


Fig. 1. Best MAE for each model across three datasets.

05 Analysis of SVR forecasts

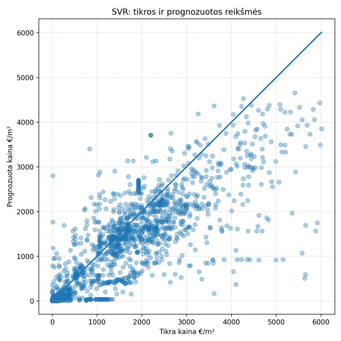


Fig. 2. Actual and predicted prices (€/m²).

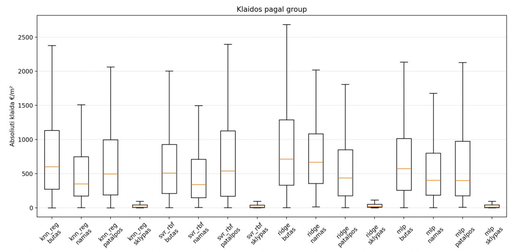


Fig. 3. Absolute errors by model and object type.

Most forecasts cluster around the line of identity; for typical properties, the margin of error is **< 1%.** The largest errors occur in the segment comprising **higher-priced, non-standard properties and detached houses**, while the smallest errors are found for land plots.

06 Allocation of the transaction price

The historical proportions method reduced the error from **€100,525 to €17,025** (test sample, 1,331 objects).

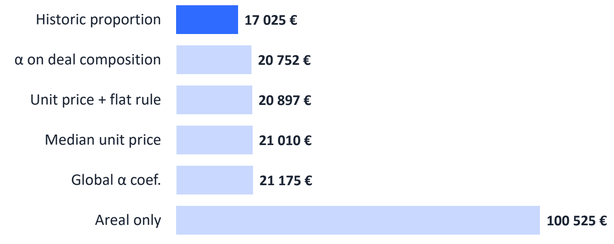


Fig. 4. MAE of price allocation methods (€).

07 Conclusions

01 **An expanded dataset** (30,682 records) has been compiled, combining transaction data and real estate listings, and incorporating a methodology for allocating transaction prices along with the automated identification of valuation zones.

02 **SVR is the most accurate model**: MAE = €382.84/m², RMSE = €706.19/m², R^2 = 0.715. However, it takes the longest to train (67 s).

03 Results depend not only on the algorithm but also on **data quality and preparation -** advertisements expand the sample size, but they also introduce noise.

04 The model is suitable as **an auxiliary tool for an initial valuation** (classifieds portals, valuation of collateral); a final decision requires an expert.

05 **In the future**: more property characteristics (condition, finishes), spatial and infrastructure data, and macroeconomic indicators.

Main references

Deppner, J. et al. (2023). Boosting the accuracy of commercial real estate appraisals. *J. Real Estate Finance and Economics.*

Numan, J. A. A., & Yusoff, I. M. (2024). Identifying the current status of real estate appraisal methods. *Real Estate Management and Valuation.*

Soegianto, L. M. et al. (2024). Comparison of model performance on housing business using LR, RF, SVR and NN. *Procedia Computer Science, 245.*