AGILE: GIScience Series, 5, 41, 2024. https://doi.org/10.5194/agile-giss-5-41-2024
Proceedings of the 27th AGILE Conference on Geographic Information Science, 4–7 Sept. 2024.
Editors: Alison Heppenstall, Mingshu Wang, Urska Demsar, Rob Lemmens, and Jing Yao.
This contribution underwent peer review based on a full paper submission.

Artificial intelligence for greater transparency in housing price estimation

Christian Mueller-Kett[1]

[1] IT & Technology, IU International University of Applied Sciences, Erfurt, Germany

Correspondence: Christian Mueller-Kett (christian.mueller-kett@iu.org)

Abstract. This paper investigates the use of machine learning (ML) models to predict housing prices. A well-performing housing price model was trained, which can seamlessly be integrated into public sector processes to increase market transparency and is based on modern ML and feature engineering methods. For these models, particular consideration was given to the spatial component. The research uses the Design Science Research approach, with a case study carried out in the city of Duisburg, Germany. The ML models developed showed better performance than traditional models. The models were embedded in official processes using Shapley values to increase interpretability. The study concludes that ML models can contribute to increased market transparency in the real estate sector.

Keywords. AI in GIS, Housing Price Valuation, Spatial Data Science

1 Introduction

A rise in real estate prices has been observed across Europe for years (Statista, 2023, Melecky & Paksi, 2023). Price expectations on the market are distorted by various factors, including "trend chases". Price estimates, often based on asking prices on internet portals rather than actual purchase contracts (e.g., Hromada, 2016; Rey-Blanco et al., 2024), encourage this phenomenon. This can lead to inflated asking prices and bubbles (Glaesner & Nathanson, 2015). The consequences are often stigmatizing neighbourhoods, segregation, and gentrification (DeSilva et al., 2012, Engelhardt, 1991, Essafi & Simon, 2015, Fuller et al., 2020, Guerrieri et al., 2013).

To counteract this, objectifying the market view by analyzing actual purchase contracts can help. For this purpose, countries use different legally regulated procedures for price estimation that are carried out by the official institutions and are usually made available to the public. For example, three categories of procedure are permitted in Germany (ImmoWertV, 2022, Maier & Herath, 2015), which can also be found in a similar form in other countries:

1. The asset value method is based on the building fabric, the production costs, and the replacement value of the property.
2. The income capitalization approach, also called value-in-use, is based on the potential income to be generated by the property.
3. The comparative value method, also called the market value, is based on actual purchase prices of similar properties and is usually used to estimate the price of residential properties.

While the comparative value was traditionally calculated manually, it is mainly based on multivariate linear regression models today. The results are usually centrally displayed by public institutions and sometimes made available to the public free of charge.

However, this method has some disadvantages: Only linear relationships are recorded in multivariate regression models. Also, numerical variables are usually classified into value range categories when using this method, adding a certain degree of subjectivity in choosing class boundaries. In addition, the spatial component, such as spatial autocorrelation, is generally not considered or is only considered insufficiently.

Furthermore, discrete feature classes and spatial reference value zones as results of these analyses are only partially in line with reality. For example, standardization of the contract year does not allow consideration of seasonality during the year.

As an alternative to this approach, more complex machine learning (ML) models are not subject to many of these disadvantages and, unlike linear regression models, can map non-linear relationships. In recent years, ML methods have been increasingly used to estimate housing prices. For example, Gu et al. (2011) used Genetic Algorithm and Support Vector Machines (G-SVM) to estimate real estate prices in China from the years 1993 to 2002, while Bahia (2013) used neural networks to predict prices in Boston, USA. These studies used individual ML models.

Banerjee & Dutta (2017), for example, went one step further, using not just one model but a support vector machine (SVM), random forest (RF), and neural networks for prediction. Phan (2018) used Stepwise Regression, Principal Component Analysis (PCA), Polynomial Regression, Regression Trees, Neural Networks, and the SVM algorithm to predict housing prices in Melbourne City, Australia, from 2016 to 2018. Mu et al. (2014) used SVM, Least Squares SVM, and Partial Least Squares for predictions in Boston, USA, while Park & Bae (2015) estimated real estate prices in Fairfax County, Virginia, USA, using the C4.5, RIPPER, Naïve Bayes, and AdaBoost algorithms.

In these studies, several ML algorithms were considered, but in a further step, Chen et al. (2017) combined an SVM with hedonic regression models to produce a joint prediction for housing prices in Taipei City, Taiwan, for the years 2007 to 2010. Truong et al. (2020) also combined RF, XGBoost, LightGBM, and hybrid regression models into a stacked model for predictions in Beijing (2009-2018). Further meta-regressors that combined the predictions of different ML models can be found in Fan et al. (2018), Lu et al. (2017), and Fourkiotis & Tsadiras (2023).

These studies present trained ML models, whereby the transfer into practical application is not the subject of discussion. Furthermore, the models developed are primarily aimed at an academic audience or private valuers. Use in official public application areas is not supported, which is also reflected in the fact that local legislation is neglected in these models. As an example related to the use case of this study in Duisburg, Germany, official valuers are not allowed by law to use the mentioned models in their official valuation procedures, as these models are not transparently interpretable and cannot be integrated into existing legal administrative processes.

Furthermore, more modern algorithms and systematic frameworks for model combinations are now available. Another identified research gap is that the spatial component is only given a subordinate role in most studies. Cellmer et al. (2020) focused more clearly on this aspect by training a geographically weighted regression model to estimate housing prices based on socio-demographic data in Poland in 2018. However, only a single ML model was considered in this study.

In addition, the results of most of these studies are difficult to transfer due to the considerable spatial and temporal distance in the data basis and the highly heterogeneous and fluctuating price trends.

Accordingly, this study aims to develop a

1. well-performing housing price model, which
2. can seamlessly be integrated into public sector processes to increase market transparency and
3. is based on modern ML and feature engineering methods, which,
4. in particular, consider the spatial component.

2 Methodology

This study applied the Design Science Research (DSR) methodology approach (Hevner et al., 2004; Hevner, 2007; Oesterle et al., 2010; Peffers et al., 2006). This approach has become established in many applied sciences in recent years and is particularly suitable for applied research in engineering and IT, which aim to produce a so-called artifact in addition to the increased knowledge base. A suitable ML model was created as the artifact of this study and the result of the design. The evaluation and diffusion (= transfer) of the artifact are based on model metrics and assessment of the integration of the model into existing official administrative processes.

In the analysis phase, the knowledge base was created through a literature review of academic journal papers, an evaluation of state-of-the-art ML and data engineering methods and frameworks, an evaluation of legal texts and guidelines for determining housing prices, and intensive and regular expert interviews.

In addition, the project environment can be described as follows: Within the framework of the DSR methodology, the expert committee on housing prices with its associated office of the City of Duisburg was the practical partner of this project. This organization has a specific scope of action within legal process specifications and structures (e.g., ImmoWertV, 2022). The technology used was R with R Markdown for automated and reproducible analysis and presentation of results, as well as documentation. The analyzed data was subject to a high degree of standardization of the data source (central purchase price collection) and data sink (BORIS data model and purchase price calculator of the state of NRW, BORIS-NRW, 2024). The people involved were employees from the office of the expert committee, some of whom had knowledge of GIS and R, a comprehensive understanding of the data, and domain-specific expertise.

2.1 Area of Investigation

A case study was carried out in the city of Duisburg to develop and evaluate the artifact. With a population of around 500,000, Duisburg is a typical medium-sized city in western Germany and, with an area of around 230 km², is one of the largest cities in the country (Statista, 2024a, 2024b), while Europe's largest inland port occupies a considerable proportion of this space (Statista, 2024c). Like many European cities of its size, Duisburg is a post-industrial city that is confronted with rising housing prices. In this study, the submarkets, "detached and semi-detached houses" and "condominiums," were considered.

2.2 Data and Software Availability

Actual real estate purchase prices from purchase contracts from 2019 to the third quarter (Q3) of 2023 were available for analysis for both submarkets. Properties included the condition, year of construction, living space area, etc. (see Tab. 1). As these data contain personally identifiable information (PII), they are restricted to administrative offices and cannot be shared publicly.

Table 1. Overview of considered object attributes per submarket.

Detached and semi-detached houses	Condominiums
- Year of contract	- Year of contract
- Standard	- Standard
- Living area	- Living area
- Rental situation	- Rental situation
- Level of location in terms of accessible infrastructure	- Level of location in terms of accessible infrastructure
- Basement availability	- Balcony availability
- Property size	- Number of units in the building
- Object status	
- Type of building	
- Subtype of building	

Due to its suitability for statistical analyses and popularity within the application domain, R was used as the analysis software in combination with several openly available libraries, such as R Markdown, Shiny, and H2O.

2.3 Data preprocessing

The data preparation consisted of the following steps:

- Imputation: In consultation with the practice partners, various simple methods for imputation were applied, such as mean and mode substitutions for applicable cases
- Spatial and temporal coverage: The purchase cases were displayed in an interactive map with regard to the spatial and temporal distribution of the purchase cases. It became clear that the data covered the entire study area satisfactorily not only in terms of time but also spatially.
- Data classification: Quantiles and natural breaks (Jenks) were used to assign numerical values to categorical classes. To further support the class formation, violin plots that reflect the distribution of the values in the respective variables were generated. The classification of the numerical features was restricted to use in linear regression models. Categorical variables were subjected to one-hot coding.
- Multicollinearity: Few collinear features were found, such as the type of building, having a balcony, and the rental situation, but these were not excluded from the analysis as these features were of technical interest for further consideration by the practice partners.

As can be seen in Tab. 2, 1579 actual purchase cases were available for the "detached and semi-detached houses" submarket and 3699 for the "condominiums" submarket once the data processing had been completed. It can also be seen that a sufficient number of samples were available for all the years under consideration.

Table 2. Overview of the number of prepared samples per year and submarket.

Years	Submarkets	
	Detached and semi-detached houses	Condominiums
2019	288	769
2020	370	812
2021	292	913
2022	408	788
2023	221	417
Total	1579	3699

2.4 Development of the Artefact

Various ML models were trained to develop the artifact. The quality of these models essentially depends on three influences:

1. The selected ML algorithm
2. The model hyperparameters
3. The input features

The H2O framework was used to select the ML algorithm systematically and for hyperparameter tuning (Fryda et

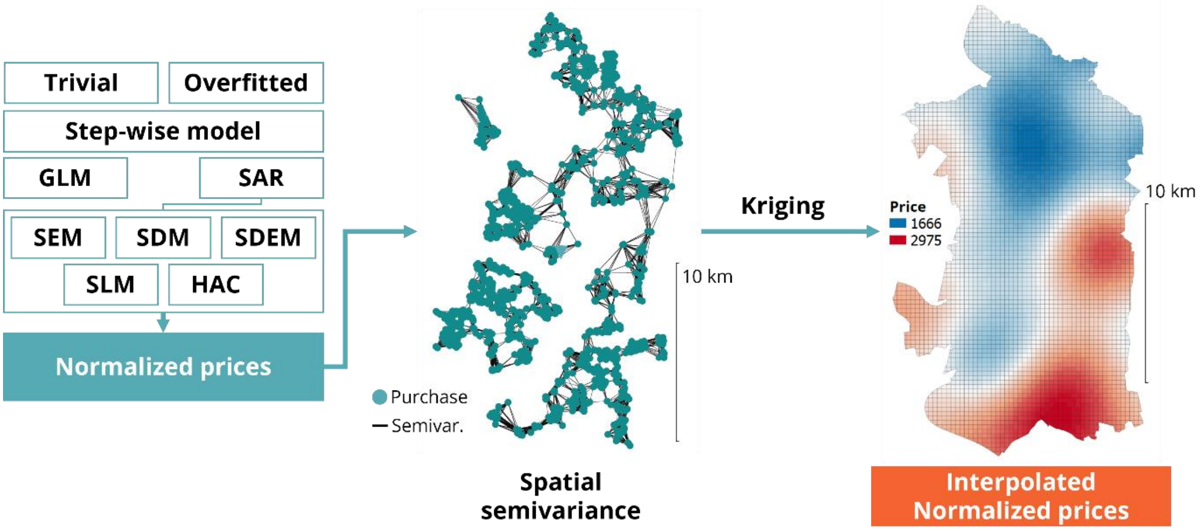

Figure 1. Process overview for generating interpolated normalized housing prices.

al., 2024). This framework considers and tests numerous ML algorithms, such as Deep Learning (Neural Networks), Distributed Random Forests (DRF), Generalized Linear Models (GLM), Generalized Additive Models (GAM), Gradient Boosting Machines (GBM), AdaBoost, SVMs, and Stacked Models, in which several models are combined into one model.

To provide the models with information that may only be available indirectly from the data, several feature engineering steps were carried out, which can be divided into three categories.

(1) To map seasonal price fluctuations during the year, various date features were generated based on the contract date of the respective purchase contracts (day of the month, day of the quarter, week of the year, quarter, month). Models considering these generated features are abbreviated to DF (= Date Features) in the following.

(2) Special attention was paid to the spatial component and its possible price-relevant influence. Consequently, a spatial cross-feature was generated, resulting from the sum of the easting and northing values of the coordinates of the properties sold. Models that used this cross feature for training are abbreviated CF (= Cross Feature) in the following.

(3) As the third and most extensive feature engineering step, normalized purchase prices were determined based on traditional regression methods and then spatially interpolated (see Fig. 1). For this purpose, a city-wide standard object was defined, and various linear regression models were then trained: A trivial model that predicted

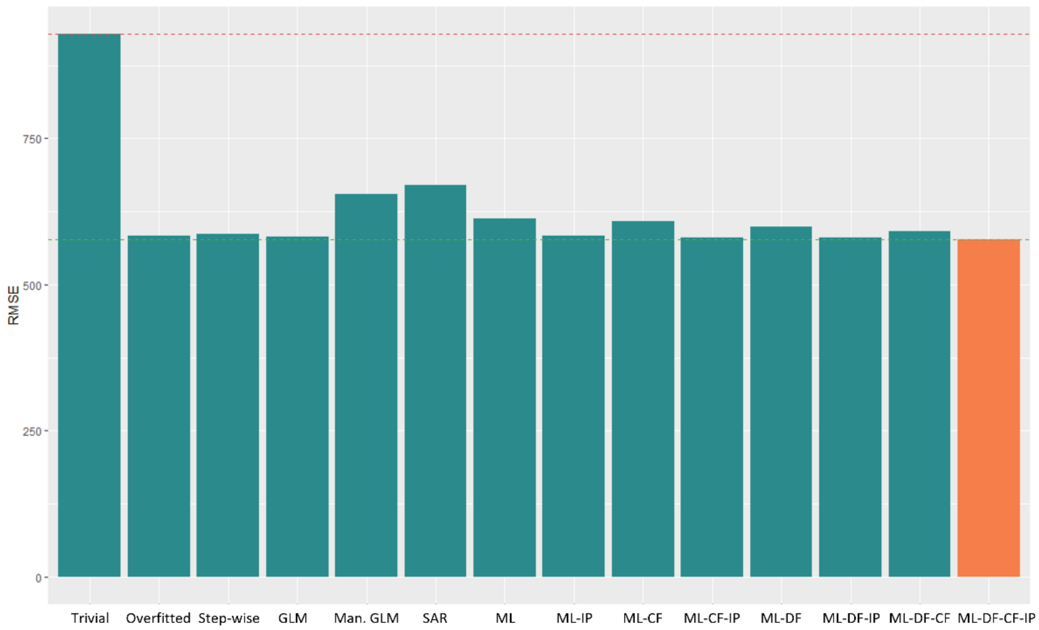

Figure 2. Model performance measured by the RMSE for the “detached and semi-detached houses” submarket.

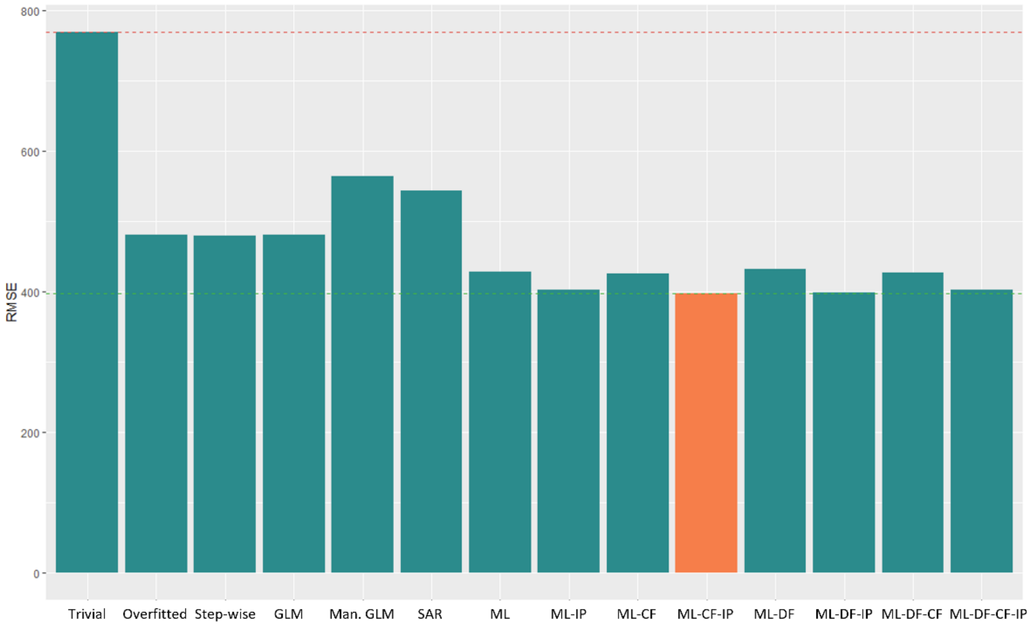

Figure 3. Model performance measured by the RMSE for the “condominium” submarket.

the arithmetic mean regardless of the characteristic value (maximum underfitting), a maximum model that took all characteristics into account, regardless of statistical significance (maximum overfitting), and a stepwise model to determine the optimum combination of features. Since the residuals showed neither homoscedasticity (Detached and semi-detached houses: BP = 169.04, p < 0.001; Condominiums: BP = 204.37, p < 0.001) nor a normal distribution (Detached and semi-detached houses: W = 0.98, p < 0.001; Condominiums: W = 0.98, p < 0.001), a non-parametric Generalized Linear Model (GLM) was also trained. As there was also a significant spatial autocorrelation (Detached and semi-detached houses: Moran’s I = 0.07, E = , -0.005, p < 0.001; Condominiums: Moran’s I = 0.232, E = , -0.004, p < 0.001), five spatial autoregressive models were trained: A Spatial Durbin Model (SDM), a Spatial Durbin Error Model (SDEM), a Spatial (y-) Lag Model (SLM), a Spatial Error Model (SEM), and a Non-parametric Heteroskedasticity and Autocorrelation Consistent (HAC) model (Pebesma & Bivand, 2023; Bivand, 2024). Geographically Weighted Regression was not applied as global regression coefficients were needed to deduct conversion coefficients for the comparative value method in the official administrative procedures. The Root Mean Squared Error (RMSE) was used as an error measure, which was also used to select the spatial model that showed the best performance. This is abbreviated as SAR (= Spatial Autoregressive Regression) in the following. By putting the regression coefficients in relation to each other to achieve relative values (= conversion coefficients) and applying these coefficients to the samples, normalized purchase prices were finally calculated. These were to be interpreted to represent the prices that would have been paid for the respective objects if they had matched the standard object in terms of their characteristics (Mann, 2005). Consequently, these normalized purchases could be directly compared, an aspect that is not considered in many studies.

In the last step, further attention was paid to the spatial component. Spatial neighbors’ relationships were calculated for the normalized prices, and a semivariogram was generated. This was used to interpolate the prices utilizing kriging. The interpolated normalized purchase prices were finally appended to the purchase cases as a newly generated feature. Models that included this feature in their training are abbreviated as IP (= Interpolated Price) below.

3 Results

The results are presented in two parts: The trained ML model as the created artifact and the embedding of the artifact in official processes.

3.1 The Trained ML Model

The ML models that showed the best performance in terms of RMSE for both submarkets were stack models. For the “detached and semi-detached houses” submarket, 7 Gradient Boosting Machines, 9 Neural Networks, a GLM, and a Distributed Random Forest were combined. The model for the “condominium” submarket combined 5 Gradient Boosting Machines, 11 Neural Networks, and two Distributed Random Forests. A GLM was used as the meta-learner for these base models for both submarkets.

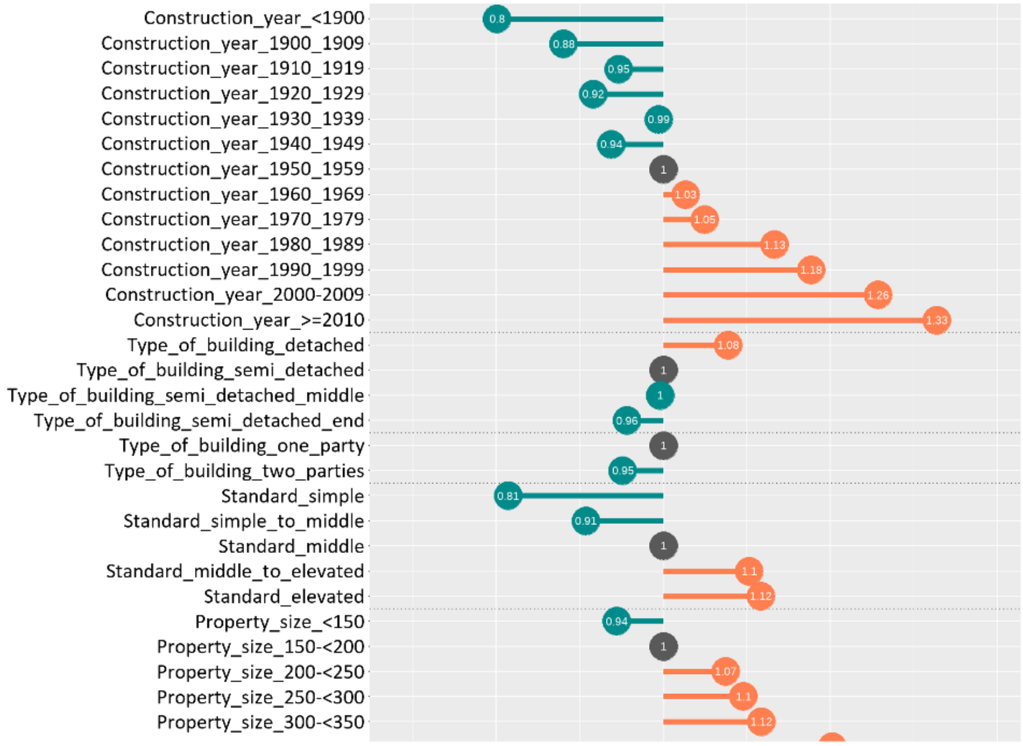

Figure 4. Sample excerpt of preliminary conversion coefficients for premiums or discounts based on property traits.

As shown in Fig. 2,3, the ML models performed better on average than the linear regression models, with this improvement particularly evident in the "condominium" submarket. The model that used all generated features (DF, CF, IP) showed the best model quality for the "detached and semi-detached houses" submarket. For the "condominiums" submarket, the model that did not use the generated date features (CF, IP) had the best predictive power.

3.2 Embedding in Official Processes

In the classic valuation method using linear regression models, conversion coefficients are derived from the

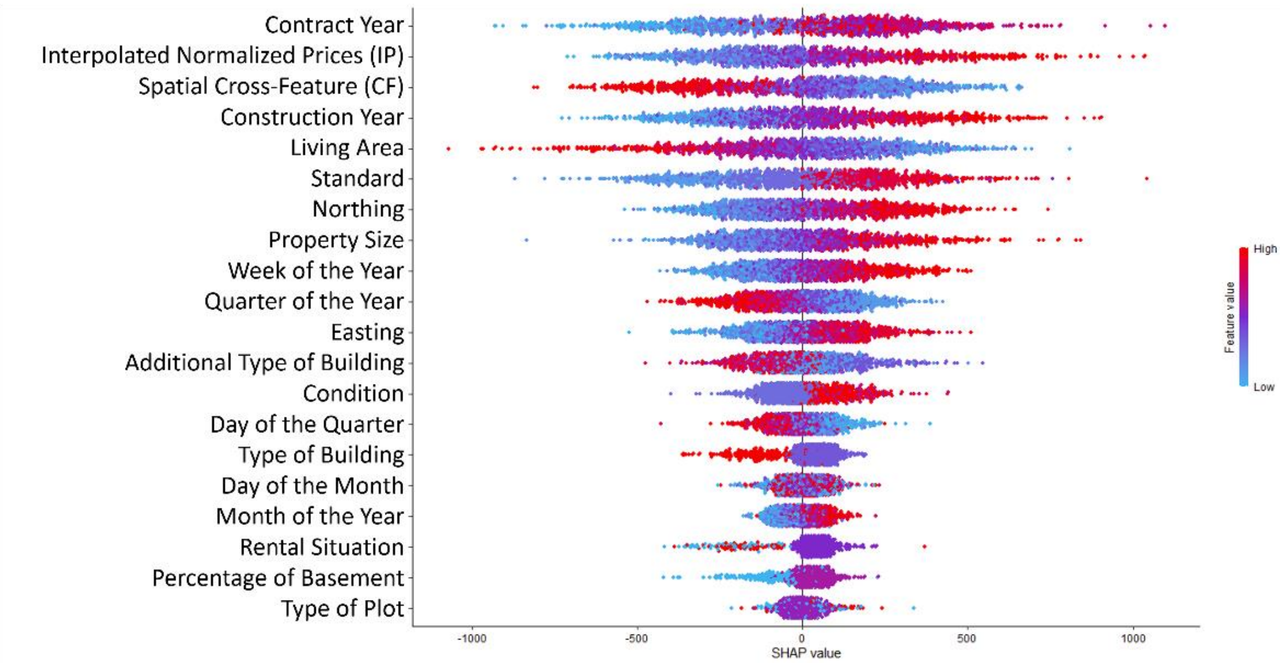

Figure 5. Shapely values to provide better interpretability of the ML models (negative/positive SHAP values decrease/increase the price; blue/red points represent low/high values of the original samples).

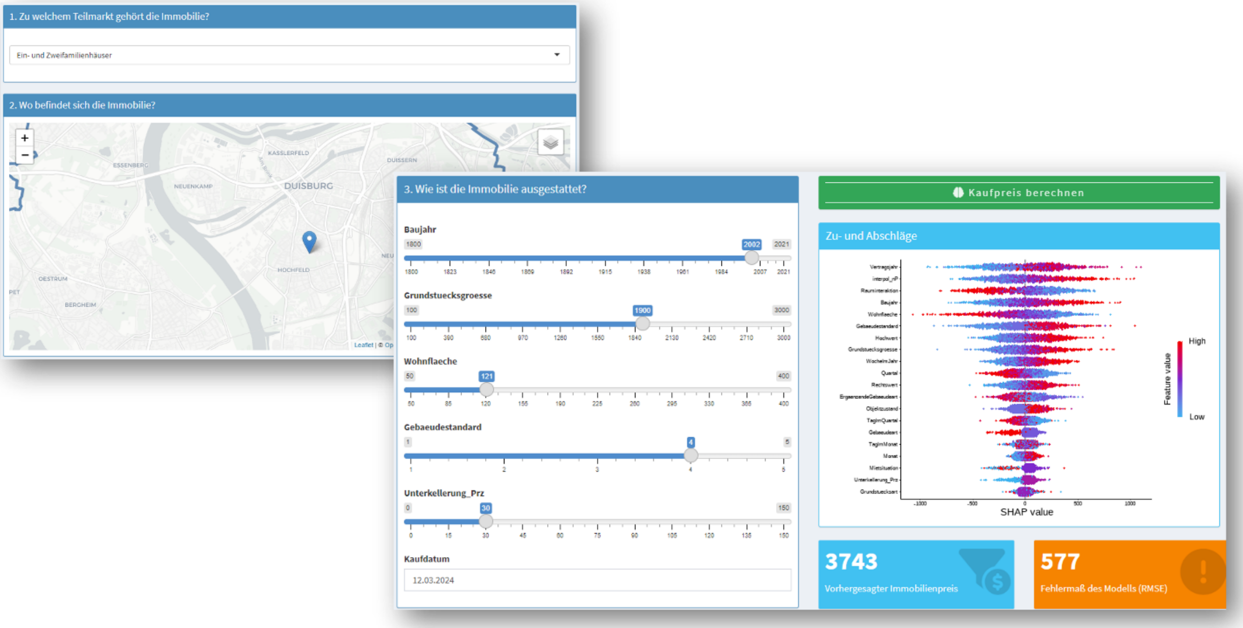

Figure 6. Illustrative example of the web interface for the interference of the model. The user is asked to specify (1) the submarket, (2) the location of the property, and (3) the traits of the object and is then presented with a model prediction and respective Shapley values.

regression coefficients. These conversion coefficients result in premiums or discounts for certain combinations of characteristics in relation to the estimated standard property value (Mann, 2005). As an example, Fig. 4 shows an excerpt from the conversion coefficients known to the experts. These coefficients support the traceability of the price determination, not least in court.

Such interpretable coefficients are not available with the stacked ML models. To increase the explainability of the predictions of the ML models, Shapley values can be used as equivalents (see Fig. 5; also see Shap, 2024). As can be seen in Fig. 5, this graphical presentation of Shapley values is comparable to the conversion coefficients in Fig. 4, which are well-known to the domain practitioners. These Shapley values, together with the housing prices predicted by the model, were displayed in a web application that official evaluators can access in a web browser in the diffusion phase of the study (see Fig. 6).

The Shiny framework for R was used for this purpose. The web application developed for model interference is similar in form and function to the applications known to users for value determination. The main differences are a more modern appearance and some helpful additional functionalities for the interpretability of the model. The similar presentation and usability of the ML models to known official purchase price calculators are supposed to promote the integration and applicability of the developed models in practice. This deployment type facilitates access to the use of the ML model and thus contributes significantly to knowledge transfer and practical applicability.

4. Discussion & Conclusion

It was shown that the trained ML models performed better than the traditional ones, especially for the "condominium" submarket. A strong orientation towards official processes and the associated data structures facilitated the transfer of the developed models into official practice. Minor deviations from these processes and structures, which can be estimated to be acceptable, allow the derivation and use of more informative features, the abandonment of discrete reference value zones and class boundaries, and, thus, a continuous and more realistic representation.

This is offset by a lower interpretability of the models and their predictions compared to the determination with easily understandable and comprehensible conversion coefficients. However, as has been shown, Explainable AI approaches, such as Shapley values, can significantly increase the interpretability of black box models and make them suitable for practical use. In contrast to the models described in the literature, this aspect of interpretability, which is comparable to the usage of conversion coefficients, together with the strong reference to existing legal requirements for data preparation, presentation of results, and accessibility of the model, contributes decisively to the usability in official valuation procedures.

Nevertheless, an evaluation of embedding in official processes cannot be conclusive. First assessments by the experts suggest that the models developed will be highly usable in official practice and thus contribute to increased market transparency. Still, it must be emphasized that no model will ever substitute official valuers but only support their work. Also, a comprehensive evaluation of this practice transfer will only become apparent in the long term and highly depends on the interpretability of the predictions. The model performance with regard to data drift and other aspects should also be monitored regularly in the future.

Finally, the DSR methodology can be seen as useful for this and similar studies in which a high value is placed on transfer to practice.

References

Bahia, I. S. H. (2013). A data mining model by using ANN for predicting real estate market: Comparative study. *International journal of intelligence science*, 3(04), 162-169

Banerjee, D., & Dutta, S. (2017). Predicting the housing price direction using machine learning techniques. In *2017 IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI)*, 2998-3000.

Bivand, R. (2024). Spatialreg [R Library], available at https://github.com/r-spatial/spatialreg/, last access: 8 February 2024

BORIS-NRW - Official information on the property market. (2024). Office of the Higher Expert Committee for Property Values in the State of North Rhine-Westphalia - Cologne District Government, , available at https://www.boris.nrw.de/boris-nrw/?lang=en, last access: 8 February 2024

Cellmer, R., Cichulska, A., & Bełej, M. (2020). Spatial analysis of housing prices and market activity with the geographically weighted regression. *ISPRS International Journal of Geo-Information*, 9(6), 380-399.

Chen, J. H., Ong, C. F., Zheng, L., Hsu, S. C. (2017). Forecasting spatial dynamics of the housing market using support vector machine. *International Journal of Strategic Property Management*, 21(3), 273-283.

DeSilva, S., & Elmelech, Y. (2012). Housing inequality in the United States: Explaining the white-minority disparities in homeownership. *Housing Studies*, 27(1), 1-26.

Engelhardt, G. V., & Poterba, J. M. (1991). House prices and demographic change: Canadian evidence. *Regional Science and Urban Economics*, 21(4), 539-546.

Essafi, Y., & Simon, A. (2015). Housing market and demography, evidence from French panel data. *European Real Estate Society*, 2015, 107-133.

Fan, C., Cui, Z., & Zhong, X. (2018). House prices prediction with machine learning algorithms. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, 6-10.

Fryda, T., LeDell, E., Gill, N. Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyoun, P., Kurka, M., Malohlava, M., Poirier, S., Wong, W., Rehak, L., Eckstrand, E., Hill, B., Vidrio, S., Jadhawani, S., Wang, A., Peck, R., Gorecki, J., Dowle, M., Tang, Y., DiPerna, L., Maurerova, V., Syzon, Y., Valenta, A., Novotny, M. (2024). R Interface for the 'H2O' Scalable Machine Learning Platform [R Library], available at https://cran.r-project.org/web/packages/h2o/h2o.pdf, last access: 8 February 2024

Fourkiotis, K. P., & Tsadiras, A. (2023). Comparing Machine Learning Techniques for House Price Prediction. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 292-303. Cham: Springer Nature Switzerland.

Fuller, G. W., Johnston, A., & Regan, A. (2020). Housing prices and wealth inequality in Western Europe. *West European Politics*, 43(2), 297-320.

Glaeser, E. L., & Nathanson, C. G. (2015). Housing bubbles. In *Handbook of regional and urban economics*, Vol. 5, 701-751. Elsevier.

Gu, J., Zhu, M., Jiang, L. (2011) Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications* 38(4), 3383-3386.

Guerrieri, V., Hartley, D., & Hurst, E. (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics*, 100, 45-60.

Hevner AR, March ST, Park J, et al. (2004) Design science in information systems research. *MIS Quarterly* 28(1), 75-105.

Hevner AR (2007) A three cycle view of design science research. *Scandinavian Journal of Information Systems* 19(2), 87-92.

Hromada, E. (2016). Real estate valuation using data mining software. *Procedia engineering*, 164, 284-291.

ImmoWertV (2022). [German] Verordnung über die Grundsätze für die Ermittlung der Verkehrswerte von

Immobilien und der für die Wertermittlung erforderlichen Daten - Ordinance on the principles for determining the market value of real estate and the data required for the valuation, available at https://www.gesetze-im-internet.de/immowertv_2022/index.html, last access: 8 February 2024

Lu, S., Li, Z., Qin, Z., Yang, X., & Goh, R. S. M. (2017). A hybrid regression technique for house prices prediction. In *2017 IEEE international conference on industrial engineering and engineering management (IEEM)*, 319-323.

Maier, G., Herath, S. (2015). Real estate valuation with hedonic price models. *Theoretical principles and practical application*. Springer Fachmedien Wiesbaden, 2015

Mann, W. (2005). [German] Die Regressionsanalyse zur Unterstützung der Anwendung des Normierungsprinzips in der Grundstücksbewertung. *Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv)*, Issue 5/2005, 283-294

Melecky, A., & Paksi, D. (2023). European Housing Prices Through the Lens of Trends. *Prague Economic Papers*, 32(5), 488-519.

Mu, J., Wu, F., & Zhang, A. (2014). Housing value forecasting based on machine learning methods. In *Abstract and Applied Analysis* (2014). Hindawi.

Oesterle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E., J. (2010). Memorandum on design-oriented information systems research. *European Journal of Information Systems* 20(1), 7-10

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert systems with applications*, 42(6), 2928-2934.

Pebesma E, Bivand R (2023). Spatial Data Science With Applications in R. Chapman & Hall, available at https://r-spatial.org/book/, last access: 8 February 2024

Pfeffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research. In *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006)*, Claremont, CA, USA, 83-106

Phan, T. D. (2018). Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia. In *2018 International conference on machine learning and data engineering (iCMLDE)*, 35-42.

Rey-Blanco, D., Zofío, J. L., & González-Arias, J. (2024). Improving hedonic housing price models by integrating optimal accessibility indices into regression and random forest analyses. *Expert Systems with Applications*, 235, 121059.

Statista. (2023). Real Estate - Europe | Statista Market Forecast,available at https://de.statista.com/outlook/fmo/immobilien/europa#wert, last access: 8 February 2024

Statista. (2024a). Duisburg - Population until 2022, available at https://de.statista.com/statistik/daten/studie/322466/umfrage/entwicklung-der-gesamtbevoelkerung-in-duisburg/, last access: 8 February 2024

Statista. (2024b). Cities with the largest surface area in Germany, available at https://de.statista.com/statistik/daten/studie/1233769/umfrage/flaeche-der-grossstaedte-deutschlands/, last access: 8 February 2024

Statista. (2024c). Largest inland ports in Europe - freight transport 2022, available at https://de.statista.com/statistik/daten/studie/1125128/umfrage/gueterbefoerderung-in-den-groessten-binnenhaefen-europas/, last access: 8 February 2024

Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques. *Procedia Computer Science*, 174, 433-442.