DOI: 10.1111/1540-6229.12397

ORIGINAL ARTICLE

WILEY

Interpretable machine learning for real estate market analysis

Felix Lorenz[1] | Jonas Willwersch[1] | Marcelo Cajias[1,2] | Franz Fuerst[3]

[1]International Real Estate Business School (IRE|BS), University of Regensburg, Regensburg, Germany

[2]PATRIZIA AG, Augsburg, Germany

[3]Department of Land Economy and Trinity Hall, University of Cambridge, Cambridge, UK

Correspondence
Jonas Willwersch, University of Regensburg, International Real Estate Business School (IRE|BS), Universitaetsstrasse 31, 93053 Regensburg, Germany.
Email: jonas.willwersch@wiwi.uni-regensburg.de

Abstract

Machine Learning (ML) excels at most predictive tasks but its complex nonparametric structure renders it less useful for inference and out-of sample predictions. This article aims to elucidate and enhance the analytical capabilities of ML in real estate through Interpretable ML (IML). Specifically, we compare a hedonic ML approach to a set of model-agnostic interpretation methods. Our results suggest that IML methods permit a peek into the black box of algorithmic decision making by showing the web of associative relationships between variables in greater resolution. In our empirical applications, we confirm that size and age are the most important rent drivers. Further analysis reveals that certain bundles of hedonic characteristics, such as large apartments in historic buildings with balconies located in affluent neighborhoods, attract higher rents than adding up the contributions of each hedonic characteristic. Building age is shown to exhibit a U-shaped pattern in that both the youngest and oldest buildings attract the highest rents. Besides revealing valuable distance decay functions for spatial variables, IML methods are also able to visualise how the strength and interactions of hedonic characteristics change over time, which investors could use to determine the types of assets that perform best at any given stage of the real estate investment cycle.

KEYWORDS
black box, hedonic modeling, interpretable machine learning, rental estimation, residential real estate

1 | INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) are rapidly gaining importance across several domains. ML algorithms have become part of everyday life (e.g., personalized web ads, mail spam filters, etc.) but there is also a growing number of cutting-edge research applications. For instance, Deepmind and its interdisciplinary research team solved one of the biggest challenges in biology with their AI-based system AlphaFold to predict how proteins fold—a problem that had been investigated for nearly 50 years (Senior et al., 2020). Further high-stake domains include arrival planning in emergency departments and cancer diagnoses in healthcare settings (Ahmad et al., 2018) or recidivism forecasting in criminal justice (Berk & Bleich, 2013).

Given the vast potential of ML, it may be surprising that their uptake in many domains is much slower than what might have been expected from a purely technical efficiency point of view. In economic data analysis, ML excels as an extension of econometric regression analysis and cluster-based categorization tasks. However, because these methods are often perceived as opaque, their so-called black box character has repeatedly been criticized. Certain use cases such as AI-based decision support of credit applications may improve and accelerate the business operations of banks but also entail a heightened risk of replicating hidden biases inherent in existing structures and decision-making patterns. Consequently, explaining the inner workings of an ML model is vital for justifying decisions and generating new insights (Adadi & Berrada, 2018).

A similar picture emerges for the application of AI and ML in the real estate industry. By treating a real estate asset as the sum of its individual characteristics, the hedonic price regression has established itself as the main approach for price and rent estimation. ML models have proven helpful in real estate hedonic modeling for predictive purposes. Nevertheless, hedonic models lack transparency and do not reveal the underlying theoretical relationships (Mullainathan & Spiess, 2017). How can this obvious weakness be overcome? One possibility is to limit the complexity of models to ensure interpretability. Lechner et al. (2020), propose a deep learning algorithm that controls an automobile using only a few artificial neurons and an algorithm that is easy to understand while maintaining robustness and functionality. Another possibility is to examine existing ML algorithms and open up their results to establish interpretability. The present study follows this path. An ML algorithm eXtreme Gradient Boosting (XGB) is used for a hedonic rent estimation in a local residential real estate market to which Interpretable ML (IML) methods are then applied. Model-agnostic tools, namely feature importance (FI), feature effects, and feature interactions are applied to illustrate how each hedonic characteristic contributes to IML estimates.

To the best of the authors' knowledge, this is the first study of IML focusing on residential rent prediction.

2 | STATE OF RESEARCH

For decades, hedonic models have formed the basis for empirically assessing prices and rents of properties based on their characteristics such as amenities or location.

According to Sirmans et al. (2005), the origins of the hedonic model do not go back to a single founding person. Court (1939) first used a hedonic procedure to determine automobile prices while Lancaster's (1966) and Rosen's (1974) applications to the real estate context spawned a burgeoning body of literature. Essays by Sheppard (1999), Malpezzi (2002), and Sirmans et al. (2005) provide an overview of the diversity and complexity of the questions that arise within hedonic research. Dubin (1988) argues that building characteristics that usually determine prices in a hedonic model can be grouped into three categories: structural, location, and neighborhood variables. Can (1992) and Stamou et al. (2017) describe them as follows: 1) Structural variables pertain to individual features of a dwelling such as its size, age or the number of rooms. 2) Location variables, often represented as distance to the central business district (CBD) or other amenities, define the geographic location. 3) Neighborhood variables capture the wider socioeconomic environment of an area such as household income, crime rates or urban design features. Location and neighborhood variables are frequently considered together, as sometimes the distinction is not evident (Can, 1992; Des Rosiers et al., 2011; Haider & Miller, 2000; Stamou et al., 2017). In the recent past, much of the focus of studies has been on the effect of these locational or neighborhood characteristics. Within this group, variables of interest stem mainly from the environmental, infrastructure and social domains. With respect to features in the immediate environment of a property, Dumm et al. (2016), Rouwendal et al. (2017), and Jauregui et al. (2019) analyze the effect of proximity to water on property price. Studies by Below et al. (2015) and Dumm et al. (2018) show the price impact of nearby subsurface conditions such as sinkholes or land erosion. Other issues, such as the influence of distance to urban green spaces (Conway et al., 2010) or the presence of air pollution (Fernández-Avilés et al., 2012), also receive attention. Considering the group of neighboring infrastructural facilities and their impact on properties, different studies emerged. Hoen et al. (2015), Hoen and Atkinson-Palombo (2016), and Wyman and Mothorpe (2018) study the effects of nearby electric facilities on property prices such as wind turbines and power lines. Availability of transportation facilities, such as of a highway and rail transit, are investigated by Chernobai et al. (2011), Li (2020), and Chin et al. (2020). According to Theisen and Emblem (2018) and Zheng et al. (2016), the possibility of an easy access to early childhood education and training in the form of nearby kindergarten or schools is also a price-determining factor of residential properties. Rarer applications of hedonics include the influence of strip clubs (Brooks et al., 2020) or the proximity to food trucks (Freybote et al., 2017). Factors in the immediate social environment can also play a role. For example, Goodwin et al. (2020) find that the presence of home ownership associations has price-determining effects and Seo (2018) documents the importance of neighborhood condition.

Hedonic models can take a parametric, semi-, or non parametric functional form. Improved computing power has allowed methods, such as ML, to complement this estimation process. While the parametric hedonic price regression approach is largely applied for inferential purposes, its potential for predictive tasks is rather limited (Pérez-Rave et al., 2019). For ML applications, inference has hardly played a role so far due to their mostly opaque algorithms. Instead, they have been mainly used for making predictions. ML algorithms, like gradient tree boosting (GTB) (Friedman, 2001), random forest regression (RFR) (Breiman, 2001a), and support vector regression (SVR) (Smola & Schölkopf, 2004), are capable of artificially learning from the underlying data and continuously improving their predictive performance. Hence, these algorithms have shown remarkable accuracy. In the real estate literature, various studies demonstrate the performance of ML algorithms and parametric hedonic models, including Lam et al. (2009) and Kontrimas and Verikas (2011) for SVR, Yoo et al. (2012), Antipov and Pokryshevskaya (2012), and Yao et al. (2018) for RFR and van Wezel et al. (2005) and Kok et al. (2017) for boosting methods such as GTB. While, Zurada et al. (2011), Mayer et al. (2019), and Ho et al. (2021) document the performance of different ML methods, Cajias et al. (2021) demonstrate how they can support the analysis of real estate markets and portfolios.

ML applications have been criticized for their black box character (McCluskey et al., 2013), as the final outcome often delivers the raw prediction results without a transparent account of how these were derived. As Mayer et al. (2019) state, ML models sacrifice comprehensibility for predictive accuracy since they capture highly complex patterns within the underlying data. Researchers face a trade-off between the *what* and the *why*. Put differently, they can either limit the inputs to understand the basic mechanisms at work better or maximize the predictive power by allowing complex but ultimately unintelligible relationships.

In general, many ML methods, such as SVR, RFR, and GTB, provide model transparency by disclosing the underlying algorithm, which can be described mathematically without further knowledge of the data. Model interpretability in terms of identifying and understanding what factors impact the final predictions seems to be the bottleneck for an overall acceptance and implementation of ML methods because narrowly defined metrics of predictive accuracy are "an incomplete description of most real-world tasks" (Doshi-Velez & Kim, 2017).

In the real estate literature, first inroads have been made for combined predictive and inferential purposes within an ML context. Pérez-Rave et al. (2019) propose a variable selection process named "incremental sample with resampling" which is tested on two datasets of property prices. They apply random forests to varying subsamples to predict property prices. Variables are identified as significant if the feature is used in the final prediction rule of the RFRs for 95% of the subsamples. The final inferential interpretation is based on a parametric hedonic model using only the ML-selected variables. Moreover, Pace and Hayunga (2020) analyze the informational content of residuals from linear, spatial hedonic regression, and ML models. After applying regression trees (RTs), they find that spatial information is still present in the residuals of ML models. Although single trees are easy to understand and their decision rule can be illustrated graphically, they show limited predictive performance and tend to be unstable due to high sensitivity to changes in the data or tuning parameter. In summary, this nascent field of research is poised to enhance the interpretability of ML models and the price impacts of hedonic characteristics.

3 | DATA

Both hedonic regression models and ML-based approaches require large data samples to be effective. Our sample comprises 52,966 observations of residential rents in Frankfurt am Main, one of the major conurbations in Germany with 5.8 million inhabitants and a sizable rental segment of nearly 80% of the residential market.

Rental data were supplied by Empirica Systeme, one of the largest German provider of real estate data, which comprises, among others, real estate listings of leading German Multiple

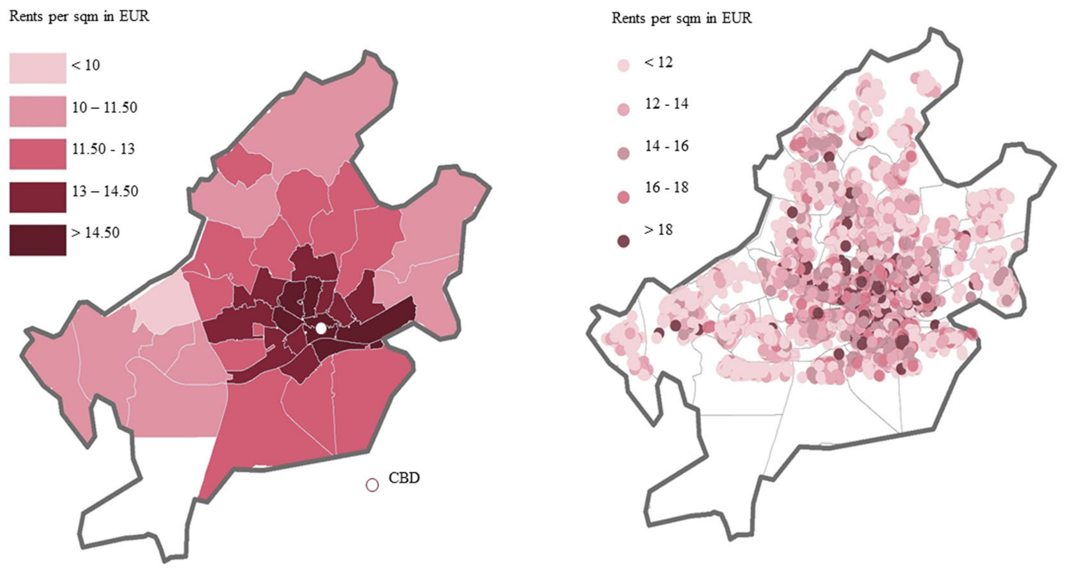


FIGURE 1 Distribution of rents and observations of the Frankfurt data sample
[Color figure can be viewed at wileyonlinelibrary.com]
Note: The left map shows average rents per square meter for each postal code. The right map depicts all observations. Both cover the Frankfurt city area from 2013 to 2019. The thin grey lines display the postcodes.
Source: Own depiction.

Listing Systems (MLS). Data preparation and cleaning are performed to account for duplicates and erroneous data points. As the study focuses on the urban rental market in Frankfurt dominated by rental apartments, we exclude all single, semidetached, and terraced houses. We, furthermore, exclude student apartments, senior living accommodations, furnished co-living spaces, and short-stay apartments to control for highly specialized submarkets that may bias the overall rental market. Figure 1 provides two maps of the rental distribution in the data sample for Frankfurt. It highlights the average rent per square meter in every postcode area (left) and displays all observations gathered (right). Both maps indicate that the highest rents are found in the center of the city with a typical leveling off gradient toward the suburban areas. There are no rental observations in the most southern part of Frankfurt due to highly forested areas and the airport of Frankfurt.

Besides the rent as a target variable, the data contain information on structural characteristics in terms of size in square meters, building age, floor, whether a kitchen, parking spot, balcony, terrace, bathtub, and elevator are present and whether an apartment is refurbished. We add socioeconomic data from Growth from Knowledge, Germany's largest market research institute. Since all rental data points are georeferenced, we are able to add a spatial gravity layer based on data from Eurostat, the German statistical office and Open Street Map to account for spatial information and, therefore, add several location variables. We include the distance to the CBD as well as to numerous important amenities.

MLS are frequently used in German rental markets by professional as well as private landlords. Moreover, since neither landlords nor tenants are obliged to disclose contractual information in Germany, listing data is the main source of information for both researchers and practitioners.[1]

[1] See Gröbel and Thomschke (2018) using German rental listing prices in research as well as well-established applications of listing data, for example, F+B Residential Index or Empirica Real Estate Index in practice.

It should also be noted that tenants are usually price takers, particularly in undersupplied urban areas.[2] A spot check of individual rentals reveals that a landlord typically receives inquiries in the double-digit range for any advertised apartment. The landlord's rental decision is not usually an auction process but is either based on creditworthiness or a 'first come, first served' rule. Cajias and Freudenreich (2018) demonstrate that German residential markets have been subject to relatively short time-on-market and diminishing degrees of overpricing over the years. Gröbel (2019) suggests that asking price data in Germany "reflect the currently prevailing overall market situation."

Table 1 shows the descriptive statistics. We find a mean asking rent of 1,036 EUR p.m. (euros per month) per apartment comprising 78 sqm on average.

4 | RESEARCH STRATEGY

ML has proven its predictive power in the literature and is commonly used by real estate professionals to inform their decision making (RICS, 2017). To show the rationale behind ML prediction, interpretation methods were established to identify how the model comes to its final decision.

4.1 | Machine learning methods and their interpretability

We apply a tree-based approach to build the foundation for further analysis. As Pace and Hayunga (2020) state, an RT is easy to understand, while still being capable of identifying a complex pattern. Trees can capture nonlinear relationships as well as interactions. At its core, RT can be understood as a set of nested if-else conditions. Tree-based models divide the data into distinct subsets and make a prediction for every subset (which is the average outcome of all observations in the specific subset). The division is made by several splitting steps, in which a feature variable is chosen iteratively and its feature space split in a way starting with the most strongly affected criterion (i.e., the one that reduces prediction error by the largest margin) until a stopping point is reached.

Since single trees are prone to misspecification, ensembles are used to aggregate and combine the prediction rule of multiple trees. We choose XGB as an ensemble boosting method that predicts property prices and rents more accurately and at the same time yields robust estimation results. Developed by Chen and Guestrin (2016), it is a promising approach for regression as well as for classification since it contains specific features that won several Kaggle competitions in the recent past.

The boosting technique fits an initial tree, calculates the residuals of the initial prediction, and fits another tree on the residuals to stepwise reduce the prediction error and incrementally enhance the final prediction rule. To prevent overfitting, cross-validation is applied. The superior performance of this algorithm is due to various factors, such as increased computational speed, the possibility of parallel processing, that is, running the code on all cores of the computer, the general boosting ability, the conversion of a weak learner into a strong learner, and the so-called regularization technique which is integrated into the algorithm to prevent overfitting. For these

[2] According to CBRE, the vacancy rate for residential real estate in the city of Frankfurt am Main is 0.4% in 2021. Moreover, Immobilienscout 24, the leading online listing platform for real estate in Germany, reports 198 clicks on average for an online apartment advertisement.

TABLE 1 Descriptive Statistics of the dataset for Frankfurt am Main (2013–2019)

	Unit	Mean	Median	Std. Dev
Rent	EUR/month	1,036.123	884	638.175
Size in sqm	sqm	78.175	72	36.688
Floors	Integer	2.396	2	2.328
Age (relative to 2017)	Integer	49.377	48	39.701
Bathtub	Binary	0.564	1	0.496
Refurbished	Binary	0.242	0	0.428
Built-in kitchen	Binary	0.688	1	0.463
Balcony	Binary	0.633	1	0.482
Parking	Binary	0.487	0	0.500
Elevator	Binary	0.449	0	0.497
Terrace	Binary	0.136	0	0.342
Purchasing Power	EUR/HH/ZIP	50,390	49,993	5,798
CBD_distance	Km.	3.616	3.604	1.896
Bar_distance	Km.	0.722	0.511	0.636
Beergarden_distance	Km.	1.135	0.937	0.759
Cafe_distance	Km.	0.346	0.240	0.325
Bakery_distance	Km.	0.370	0.245	0.403
Convenience store_distance	Km.	0.849	0.589	0.748
Department store_distance	Km.	1.550	1.306	0.997
Supermarket_distance	Km.	0.252	0.223	0.167
Bus station_distance	Km.	3.062	2.667	1.566
Railway station_distance	Km.	0.835	0.581	0.685
Traffic signals_distance	Km.	0.186	0.157	0.135
Car wash_distance	Km.	1.266	1.234	0.584
Park_distance	Km.	0.266	0.236	0.158
School_distance	Km.	0.302	0.278	0.167

Note: The table reports the summary statistics comprising data as of January 2013 to December 2019. Age is calculated as the difference of the building age to the year 2017. All distance variables are calculated as the distance to the specific dwelling in kilometers. Binary variables report whether the dwelling includes a certain characteristic (1) or not (0). Rent is presented as euro per month. Information on households (HH) is reported on ZIP level. Min: minimum value; Max: maximum value; SD: standard deviation.

and other reasons, XG Boost is frequently preferred to methods such as SVR, Gradient Boosting, or RFR, which enjoyed great popularity in the past and continue to be used.

As the internal logic and consequently the rationale behind the individual predictions remains opaque, the use of ML often lacks transparency. A growing body of literature on IML[3] has evolved in recent years to corroborate algorithmic decision making (Adadi & Berrada, 2018; Arrieta et al., 2020 or Linardatos et al., 2021; Carvalho et al., 2019). In general, tree-based ML methods show at least moderate algorithmic transparency since their underlying concept and theory is comprehensible and mathematically described (James et al., 2013). Despite these advances, it is not evident which feature contributes to the prediction to what extent.

[3] In the context of IML, the term Explainable Artificial Intelligence (XAI) is often used synonymously.

One possibility to understand how predictions are achieved in this context is to use **interpretable ML models**. They are also referred to as transparent models since they are considered to be self-explanatory. Like parametric models, restrictions can limit the complexity of the model and, therefore, allow inferential insights. RTs are a well-known example of interpretable ML models if the depth of the tree is limited. As Molnar (2020) states, short trees with a depth up to three splits are interpretable in a comprehensive way with no more than three if-else conditions sufficiently explaining how the model yields a certain prediction.

Limiting the model's complexity often results in depriving ML of its effect since their flexible structure enables a strong predictive performance (Breiman, 2001b).[4] Post-hoc and model-agnostic interpretation methods have been developed which separate the explanatory framework and the ML model, thus, preserving its predictive capabilities. In contrast to interpretable models, the ML model remains a black box, with the separated interpretation methods aiming at extracting interpretable information post-hoc. Model-agnostic tools benefit from their flexibility because they do not depend on a specific ML method and can be applied to various learners (Ribeiro et al., 2016).

Interpretation methods differ on whether they focus on FI or feature effects. The first one aims at evaluating which feature contributes the most to the prediction, whereas the second sheds light on how a single feature contributes to the prediction. The methods are perceived as useful tools for showing the impact of features in ML models and explaining the inner workings at on a global level (Hastie et al., 2009). We use the FeatureEffect, FeatureImp, and Interaction functions implemented in the iml package in R (R Core Team, 2020).

4.2 | Feature importance of the hedonic characteristics

FI measures the relevance of a single feature for the prediction. The importance of a feature is calculated by permutation of the observed feature values and its effect on the prediction error, keeping all other features constant. Based on the concept of Breiman (2001a) for random forests, Fisher et al. (2019) provides a model-agnostic framework for measuring the covariates contribution to the accuracy of an ML model called "model reliance."

Let X be the feature matrix, Y the dependent variable, and f the ML model, with the prediction error e being measured by a loss function $L(Y, f(X))$. The FI is defined as the ratio of the model error after permutation to the original model error before switching features.

$$FI\ (f) = \frac{e_{perm}\,(f)}{e_{orig}\,(f)} \tag{1}$$

The permutated error is thereby calculated as the expected error of the ML model based on the permuted feature matrix X_{perm}.

$$e_{perm}\ (f) = \mathrm{E}L\left(Y, f\left(X_{perm}\right)\right). \tag{2}$$

To visualize the most important features, every variable is ranked and plotted according to their FI. Alternatively, the FI score can also be calculated as the difference of both errors, although the ratio provides the advantage of higher comparability. We use the Mean Absolute Error (MAE) as

[4] See Shmueli (2010) for further discussion on the trade-off between model accuracy and interpretability.

loss function. By switching the feature values of all observations (e.g., an observation with 1 for a kitchen being present is switched to 0), FI calculates how much this change leads to an observable decrease in prediction accuracy. It can consequently identify whether the specific feature contributes to the overall prediction or whether its change does not perceptibly affect the outcome. Finally, we average the importance measures over 100 repeated permutations. As Fisher et al. (2019) state, FI is a helpful tool to identify influential features and increase the transparency of black box models.

Permutation FI may present biased results when features are highly correlated. In some cases, permutation of features would assume unrealistic combinations of features and, therefore, could show misleading results. Furthermore, the overall importance of two variables can be split and, therefore, underestimate the single features importance. Recent studies show first approaches to cope with this (see e.g., Molnar, 2020). We use hierarchical clustering to analyze the effect of correlated features on the interpretability of FI measures. Features are clusters based on correlation patterns. We keep the most important variable from every cluster and estimate new FI measures to identify if the importance of certain features changes when leaving out highly correlated ones. Large deviations would indicate that the importance of single features could be split between several included variables, reducing its true importance for the overall estimation.

4.3 | Feature effects of the hedonic characteristics

In addition to individual importance, **feature effects** show how a single feature influences the predicted outcome of an ML model. After the training process, an ML model has learned a specific relationship between the covariates and the target variable that can be analyzed. Partial Dependence (PD) plots visualize the marginal effects of features on the model's prediction (Friedman, 2001). The plots are based on PD functions that highlight the effect of one feature on the target variable when the average effects of all other features are accounted for. PD plots reveal useful information, that is, whether the relationship can be explained linearly or in a more complex manner.

Let once again X_j be the vector of the j variables and n be the number of observations. The PD is the effect of features of a subset X_S by marginalizing over all other features in the complement subset X_C (Zhao & Hastie, 2021). Given the ML model f, the partial function f_{x_S} is defined as:

$$f_{x_S}\ (x_S) = E_{x_C}\ \left[f\left(x_S, x_C\right)\right] = \int f\left(x_S, x_C\right)\ d\mathbb{P}\left(x_C\right) \tag{3}$$

With $d\mathbb{P}(x)$ being the marginal distribution of X_C. Marginalizing over all other features leads to a function that is solely dependent on the features X_S to be analyzed. The partial function f_{x_S} is estimated using the Monte Carlo method to average over actual features values $x_C^{(i)}$ while keeping X_S constant:

$$f_{x_S}\ (x_S) = \frac{1}{n}\sum_{i=1}^{n} f\left(x_S, x_C^{(i)}\right) \tag{4}$$

As shown in Greenwell (2017), all values of feature x_S (e.g., size) are in a first step replaced with the particular feature value (e.g., of the first observations). The ML model predicts expected output values for the newly created dataset (where all observations have the same constant feature value

x_S). Averaging over these predictions calculates the marginal effect at the particular feature value. This step is repeated n times to obtain a marginal effect for all observed feature values. Finally, the single feature values are plotted against the resulting f_{x_S}. For a linear hedonic model, based on ordinary least squares (OLS), a PD plot would show a straight line representing the specific estimated coefficient. As Zhao and Hastie (2021) state, PD plots are a valuable visualization tool for interpreting how the prediction of ML models depends on specific features.

PD plots can be misleading if features are highly correlated as their computation may be based on potentially unrealistic combinations of feature values for which the average prediction is calculated. This can bias the visualization of feature effects and lead to unreliable results. Apley and Zhu (2020) provide Accumulated Local Effect (ALE) plots that provide unbiased estimates by handling correlation of features which is a notable factor in real estate data. Instead of marginalizing over all features, conditional distributions are used to account for correlated instances. By using the difference of predictions instead of the average leads us to the pure effect that is solely based on the specific variable of interest.

$$\begin{aligned} f_{x_S,\,ALE}\ (x_S) &= \int E_{x_C|x_S}\left[f\left(x_S, x_C\right)|x_S = z_S\right]\, dz_S\ - constant \\ &= \int\int f\left(z_S, x_C\right) d\mathbb{P}\left(x_C|z_S\right)\, dx_c dz_S - constant. \end{aligned} \tag{5}$$

ALE plots are computed by first estimating uncentered local effects. By dividing the feature values in several neighborhood $N_j(k)$ which are defined by grid values z_S, we obtain local effects and avoid unrealistic combinations of feature values.

$$\tilde{f}_{j,ALE}\ (x) = \sum_{N_j(k)}^{k_j(x)} \frac{1}{N_j\ (k)} \sum_{N_j(k)} \left[f\left(z_{k,j}, x_j^{(i)}\right) - f\left(z_{k-1,j}, x_j^{(i)}\right)\right], \tag{6}$$

$$f_{j,ALE}\ (x) = \tilde{f}_{j,ALE}\ (x) - \frac{1}{\mathrm{n}} \sum_{i\,=\,1}^{n} \tilde{f}_{j,ALE}\left(x_j^{(i)}\right). \tag{7}$$

Replacing the feature values with the adjacent grid values results in the difference of predictions conditional on the feature of interest. Accumulating the effect over all neighborhoods and centring the estimates by the average prediction leads to the final ALE plots.

5 | ECONOMETRIC RESULTS

Our comparison of model performance and results starts with parametric and semiparametric models, followed by ML estimations of ML and model-agnostic IML models. The latter include a variety of FI and feature effect analyses.

5.1 | Parametric and semiparametric models

To ensure basic hedonic functionality of a hedonic rent estimation, we first apply linear, spatial, and nonlinear methods. A brief description of the methodolgy can be found in Appendix A. Table 2

TABLE 2 Results of the OLS, GAM, and SAR estimation

	Dependent variable: log Rent per month					
	OLS		GAM		SAR	
log size in sq.m.	0.939***	(0.002)	0.900***		0.928 ***	(0,008)
Floors	0.002***	(0.0004)	0.003***	(0.0003)	0.003 ***	(0,002)
Age (relative to 2017)	−0.0002***	(0.00003)	s 8.000***		−0.000 ***	(0,0001)
Bathtub	−0.032***	(0.002)	−0.016***	(0.001)	−0.032 ***	(0,006)
Refurbished	−0.015***	(0.002)	0.005***	(0.002)	−0.013 ***	(0,007)
Built-in kitchen	0.084***	(0.002)	0.077***	(0.002)	0.077 ***	(0,007)
Balcony	0.011***	(0.002)	0.025***	(0.002)	0.012 ***	(0,007)
Parking	0.053***	(0.002)	0.032***	(0.002)	0.048 ***	(0,008)
Elevator	0.053***	(0.002)	0.020***	(0.002)	0.048 ***	(0,009)
Terrace	0.041***	(0.002)	0.020***	(0.002)	0.041 ***	(0,009)
log Purchasing Power	0.406***	(0.011)	0.069***	(0.002)	0.313 ***	(0,040)
CBD_distance	−0.019***	(0.001)	s 8.692***		−0.014 ***	(0,002)
Bar_distance	−0.031***	(0.002)	s 8.579***		−0.024 ***	(0,008)
Beergarden_distance	−0.020***	(0.002)	s 8.631***		−0.015 ***	(0,005)
Cafe_distance	−0.014***	(0.003)	s 8.700***		−0.011 ***	(0,010)
Bakery_distance	−0.011***	(0.003)	s 8.842***		−0.016 ***	(0,009)
Convenience store_distance	−0.036***	(0.002)	s 8.144***		−0.035 ***	(0,007)
Department store_distance	−0.006***	(0.001)	s 8.580***		−0.008 ***	(0,005)
Supermarket_distance	−0.018***	(0.006)	s 6.487***		−0.029 ***	(0,020)
Bus station_distance	−0.028***	(0.001)	s 8.794***		−0.017 ***	(0,004)
Railway station_distance	−0.020***	(0.002)	s 8.757***		−0.020 ***	(0,007)
Traffic signals_distance	0.086***	(0.007)	s 8.243***		0.075 ***	(0,024)
Car wash_distance	0.012***	(0.002)	s 8.763***		0.007***	(0.006)
Park_distance	−0.024***	(0.006)	s 8.343***		−0.019***	(0.020)
School_distance	−0.003***	(0.005)	s 8.412***		0.008 ***	(0,006)
Constant	−34.043***	(3.087)	2.405***	(0.100)	−22.860***	(11,359)
rho					0.131***	
Time controls	Yes		Yes		Yes	
Locational controls	Yes		Yes		Yes	
Observations	52,966		52,966		52,966	
R^2	0.880				0.885	
adjusted R^2	0.880		0.898			
UBRE			0.028			

Note: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$, standard errors are displayed in parentheses. The GAM column reports the estimated degrees of freedom of the smooth terms (s) as well as their joint significance. Time controls (year and month) as well as location controls (apartment coordinates) are included in all models.

shows the results of the OLS, GAM, and spatial autoregressive regression (SAR) models. OLS and SAR display similar coefficients for all variables.

Variables that exhibit nonlinear behavior are included as splines in the GAM with the estimated degrees of freedom of smoothing terms being provided (denoted with "s"). Regarding the goodness of fit, OLS shows an adjusted R^2 of 0.880 and the SAR a R^2 of 0.885. The GAM has a

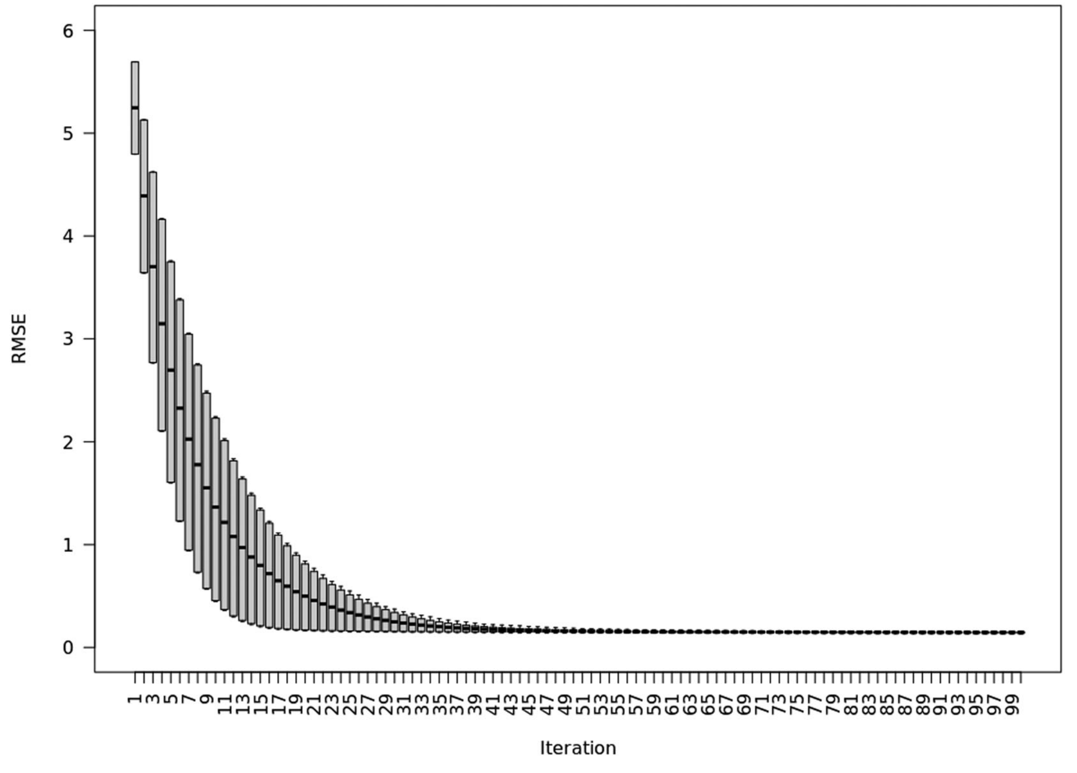


FIGURE 2 Hyperparameter tuning and model accuracy
Note: The figure displays the root mean square error (RMSE) decline per iteration step. The stacked bars represent boxplots with the black line indicating the median RMSE value.
Source: Own depiction.

marginally better fit of 0.898, which is likely due to its capability of handling nonlinear relationships. The structural hedonic variables, such as size, floor, and age, are highly significant and show expected signs. Similar findings are reported by Dumm et al. (2016, 2018) and Seo (2018) for property prices. The same holds true for the binary variables (kitchen, balcony, parking, elevator, and terrace). Gröbel and Thomschke (2018) or Stamou et al. (2017) provide analogous results for rents and prices, respectively. The location variables are also highly significant and show mostly negative signs, indicating that the smaller the distance to a certain amenity, the higher the rent. Only distance to car wash facilities and distance to traffic signals are positive, with greater distance resulting in higher rental values confirming that negative externalities associated with these features such as noise and air pollution result in a desire to not live in their immediate vicinity. In summary, the results presented in Table 2 show that the hedonic approach seems to work, and thus, justifies a more in-depth ML analysis.

5.2 | Nonparametric models and machine learning

To set up a functional ML framework, we first train the XGB algorithm on our dataset of rental prices described in the data section. We apply random cross-validation with five folds and five repetitions. Appendix B provides additional information on parameter tuning. The final XGB model

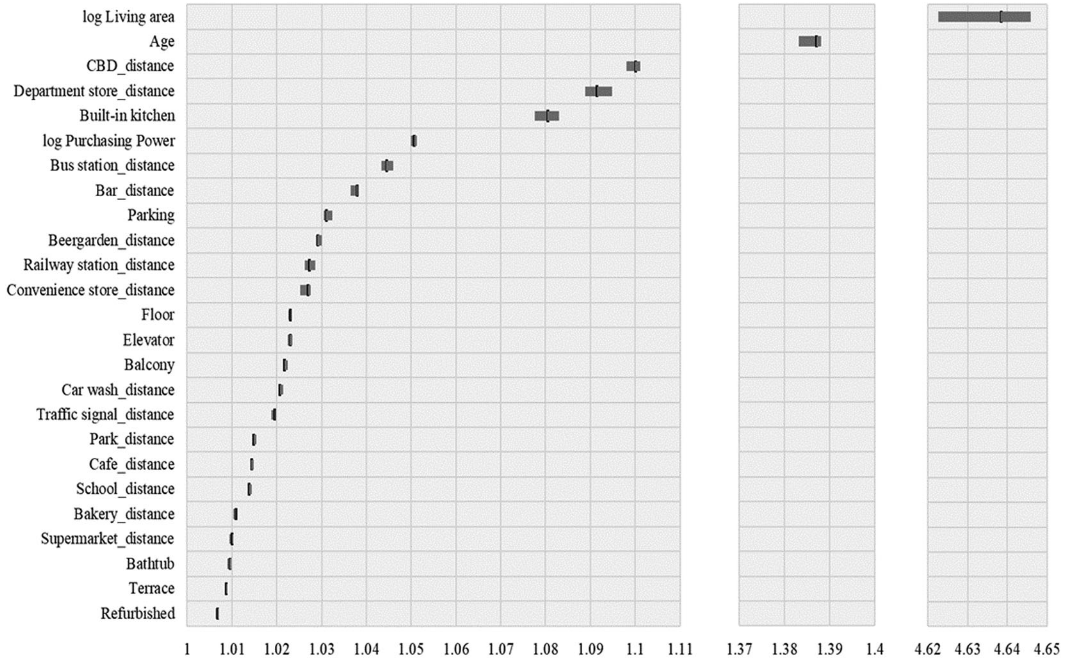


FIGURE 3 Feature importance of the hedonic characteristics

Note: The figure displays the median values of the relative feature importance obtained with XGB. MAE is chosen as loss function. Variables are ranked based on their FI score. The bar denotes the 5 and 95% quantiles of the distribution of FI scores after 100 repetitions. A break in the horizontal axis is conducted to ease readability. Time (dummies) and location controls (longitude/longitude) are included in the underlying calculation but not displayed in the figure due to little informative content. *Source*: Own depiction.

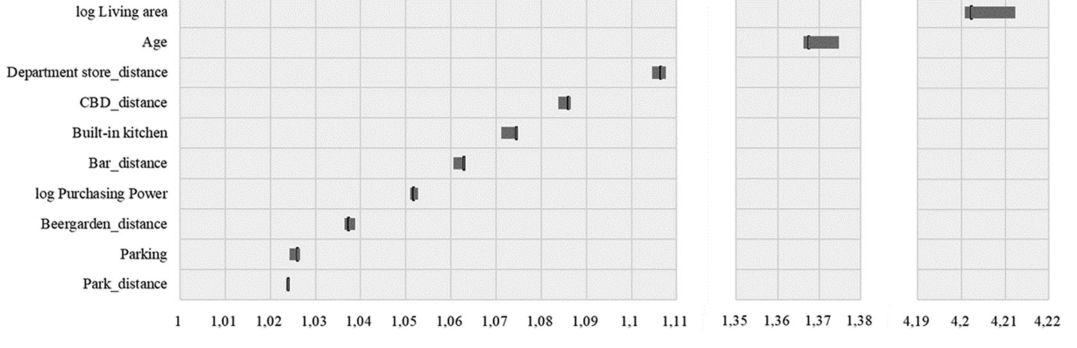


FIGURE 4 Feature importance of the hedonic characteristics based on clustering

Note: The figure displays the median values of the relative feature importance obtained with XGB. MAE is chosen as loss function. Variables are ranked based on their FI score. The bar denotes the 5 and 95% quantiles of the distribution of FI scores after 100 repetitions. A break in the horizontal axis is conducted to ease readability. Time (dummies) and location controls (longitude/longitude) are included in the underlying calculation but not displayed in the figure due to little informative content. *Source*: Own depiction.

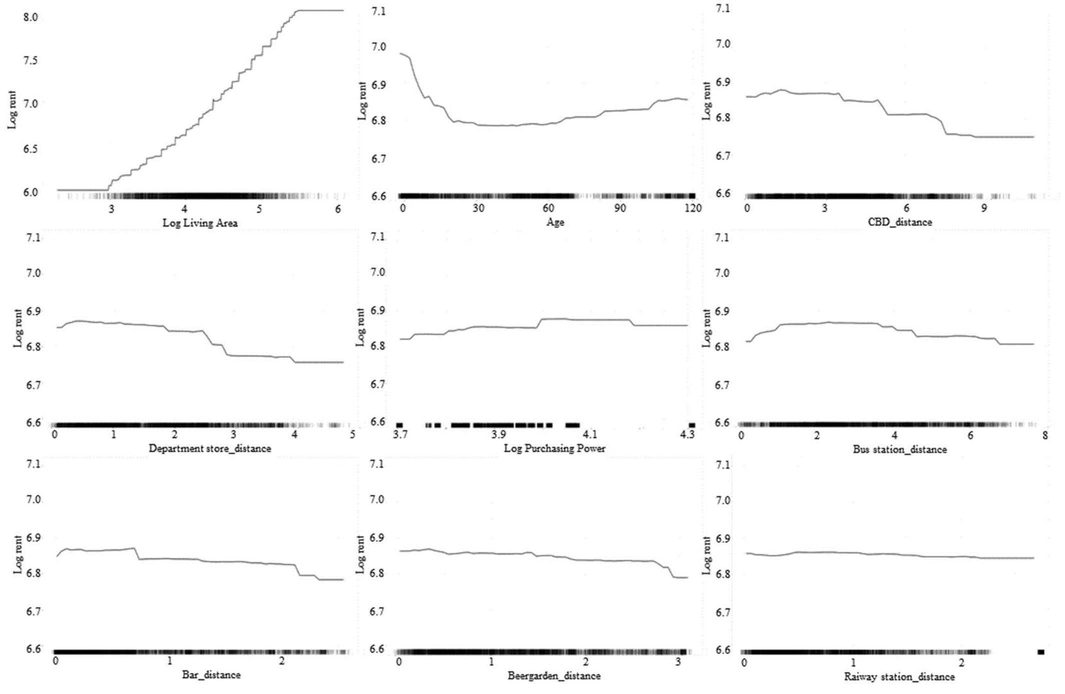


FIGURE 5 First set of PD Plots for the year 2019
Note: The figure displays the partial dependence of size in square meter, age, distance to CBD, department store, bus station, bar, beergarden, railway station, and purchasing power. The vertical axis denotes the feature values of log rent level while the horizontal axis represents the covariates feature values. Stacked black lines display the number of observations.
Source: Own depiction.

is trained with a learning rate $\eta = 0.243$ and regularization parameter of $\gamma = 0.0431$, $\lambda = 28.99$, and $\alpha = 22.64$. We found the optimal tree depth to be 38. The tuning process is graphically visualized in Figure 2. Using random search for hyperparameter tuning, we find the increase in model accuracy to go down after approximately 40 iterations.[5]

The out of sample rental prediction with XGB yields a R^2 of 92.50%. Compared to the results of the linear and nonlinear regressions presented in Table 2, XGB provides an increase of the model fit of up to 4.50 percentage points. The mean absolute percentage error marks 11.13%. Moreover, 57.96% of all predictions deviate less than 10% from the observed values. The tuned XGB algorithm subsequently allows a post-hoc analysis with a set of model-agnostic interpretation tools to identify FI and feature effects.

5.3 | Feature importance of the hedonic characteristics

Figure 3 provides the relevance of all characteristics for the ML prediction based on FI. The features are individually ranked on the y-axis from most important at the top to least important at the bottom. The x-axis provides information of how much prediction accuracy changes when the feature values are permutated. Median values are plotted with the bar denoting the 5 and 95%

[5] The tuning process takes approximately 16 hours with 72 central processing units running simultaneously.

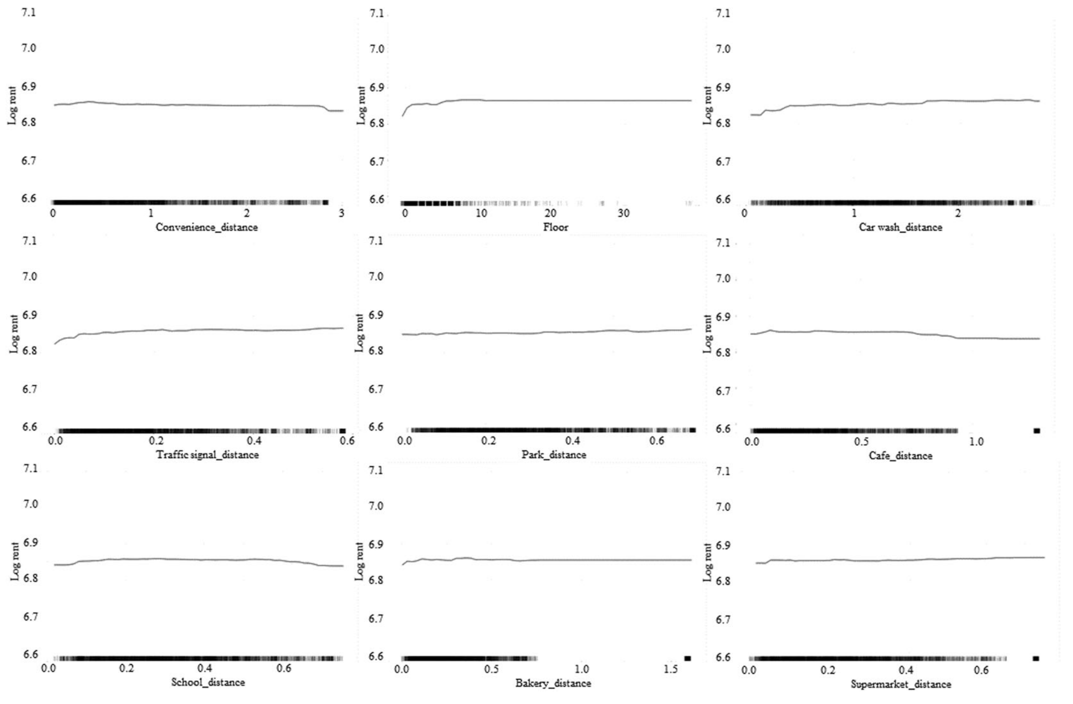


FIGURE 6 Second set of PD Plots for the year 2019
Note: The figure displays the partial dependence of the variables Distance to convenience store, floor, distance to car wash, traffic signal, park, café, school, bakery, and supermarket. The vertical axis denotes the feature values of log rent level while the horizontal axis represents the covariates feature values. Stacked black lines display the number of observations.
Source: Own depiction.

quantiles. FI ratios exceeding 1 indicate an observable impact on the overall prediction. Ratios that tend toward 1 imply a rather negligible influence of the features.[6]

It is not surprising that size and age are seen to have by far the biggest impact on rental prediction. Their median values highlight that randomly permuting size and age individually 100 times, increases the model error by a factor of 4.64 and 1.39, while keeping all other variables constant. Furthermore, the distance to the CBD and to a department store are of high importance and associated with an increase in MAE of 1.10 and 1.09. We expect both variables to be a suitable proxy for location.[7] Moreover, the presence of a built-in kitchen is also found to be influential. Purchasing power per household is followed by distance to a bus station and the nearest hospitality institutions. The existence of a parking spot complements the ten most influential variables. We will not discuss the remaining variables in detail since their contribution is rather marginal. The small distribution of FI for all variables demonstrated by the 5 and 95% quantile indicates that the results are stable over all repetitions.

[6] We calculated feature importance based on training data. There are arguments for and against the use for training data, just as there are for test data. Ultimately, it depends on whether one runs the analysis to find out how much the model prediction depends on each feature (by using training data) or how much the features contribute to the estimate when using unknown (test) data. More information about this discussion can be found in Molnar (2020).

[7] In major German cities, department stores are usually located either close to the city center or in highly frequented and, therefore, good shopping locations.

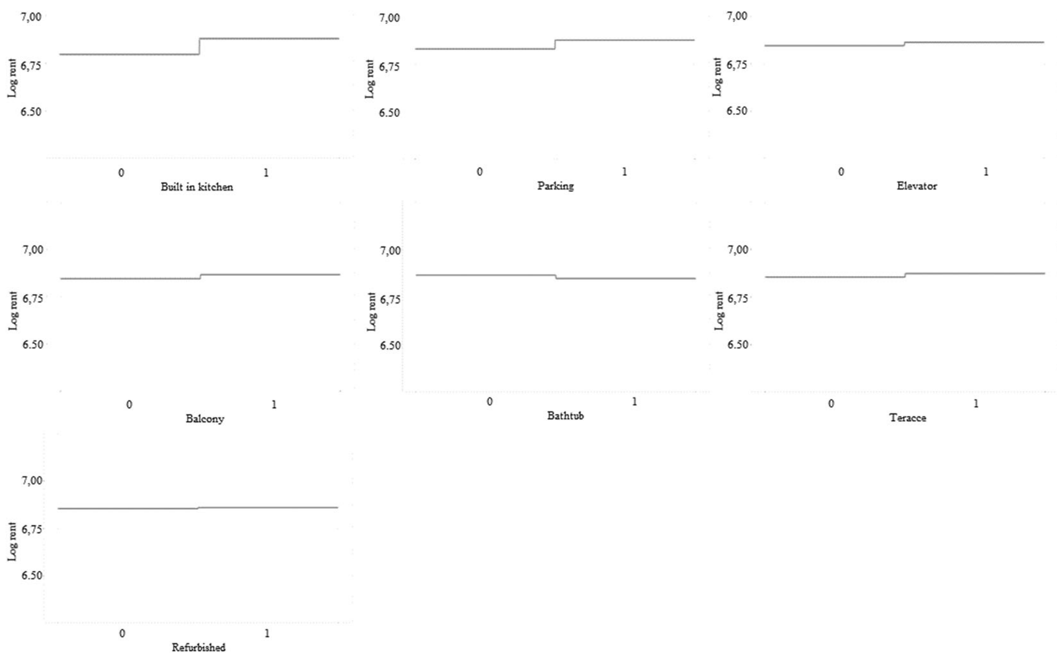


FIGURE 7 PD plots for all binary variables for the year 2019

Note: The figure displays the partial dependence of all binary variables. The vertical axis denotes the feature values of log rent level while the horizontal axis represents the covariates feature values. The distribution of the feature values can be found in the descriptive statistics in Table 1.
Source: Own depiction.

FI can show spurious results when features are highly correlated. Since the overall importance of several features combined could be split and consequently the importance of a single feature underestimated. To check whether feature correlation biases the overall interpretation of FI measures, we use hierarchical clustering. The features are, therefore, clustered based on their correlations. We keep one feature from each cluster and calculate the FI based on the subset of variables. Figure 4 shows the results. Although most variables remain stable, the importance of distance to department stores and to local bars increase significantly. Hence, we expect the influence of distance to relevant amenities to be split over several features, resulting in higher importance for single distance characteristics when leaving out correlated ones. Consequently, interpretation methods need to be handled carefully when it comes to correlation of hedonic characteristics.

To summarize, FI ranks how relevant a variable is for the predictive task as it provides which variables are more or less influential for an ML model. One can, thus, obtain a first impression whether an algorithmic hedonic model delivers reliable results that are based on a plausible understanding of the economic context. However, FI does not provide any information about the sign. To clarify whether a small or large distance is decisive, we investigate feature effects in a next step.

5.4 | Feature effects of the hedonic characteristics

PD plots enable an analysis of how any given feature influences the rental prediction and which relationships between residential rents and property characteristics has been traced by the

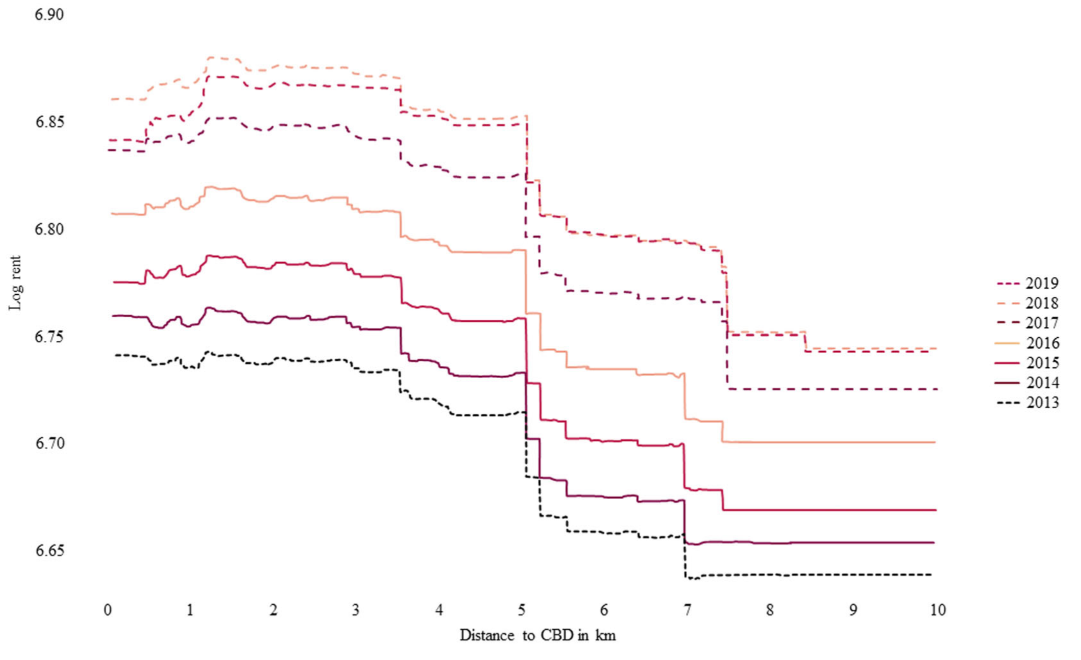


FIGURE 8 PD plot for rent and distance to CBD for the years 2013 to 2019
[Color figure can be viewed at wileyonlinelibrary.com]
Note: The figure displays the partial dependence of important variables over different periods. The vertical axis denotes the feature values of the log rent level while the horizontal axis denotes the covariates feature values.
Source: Own depiction.

algorithm. While the *x*-axis provides information on the independent variable with the stacked black lines indicating the number of observations, the y-axis shows the respective rent level. Since marginal effects are calculated and averaged for every feature value, PD plots require high computational power. The PD plot for the year 2019 illustrates this point. Figure 5 demonstrates how rental prices are associated with nine influential characteristics.

The most important feature, size in square meters, is incorporated as the natural logarithm. Since the PD plot highlights a linear relationship, the commonly applied log-log transformation can be confirmed as a good approximation of the positive relationship between apartment size and rent. Recent hedonic literature on property prices provides similar findings for the positive relationship (Dumm et al., 2016, 2018 or Stamou et al., 2017). Age is perceived to be more complex, though intuitive. We find rental values to decrease with greater age until a building year of 1990–2000. While newly build apartments obtain the highest rents, there is a clear trend toward higher depreciation in the first few years, which together with changes in living preferences as well as increasing requirements on energy-efficient construction may result in a relatively steeper decline in rental values. This is followed by a plateau of rental values up to 1940. Frankfurt was heavily bombarded in World War II, with emergency re-construction of social housing provided by the German government in the following decades. Therefore, historical pre-war buildings attract higher rents. In general, building age displays a u-shaped relationship, as reported by Mayer et al. (2019).

Distance to CBD is perceived to be highly influential. In general, we find rental prices to decline with greater distance to the city center. The hedonic literature suggests similar conclusions since

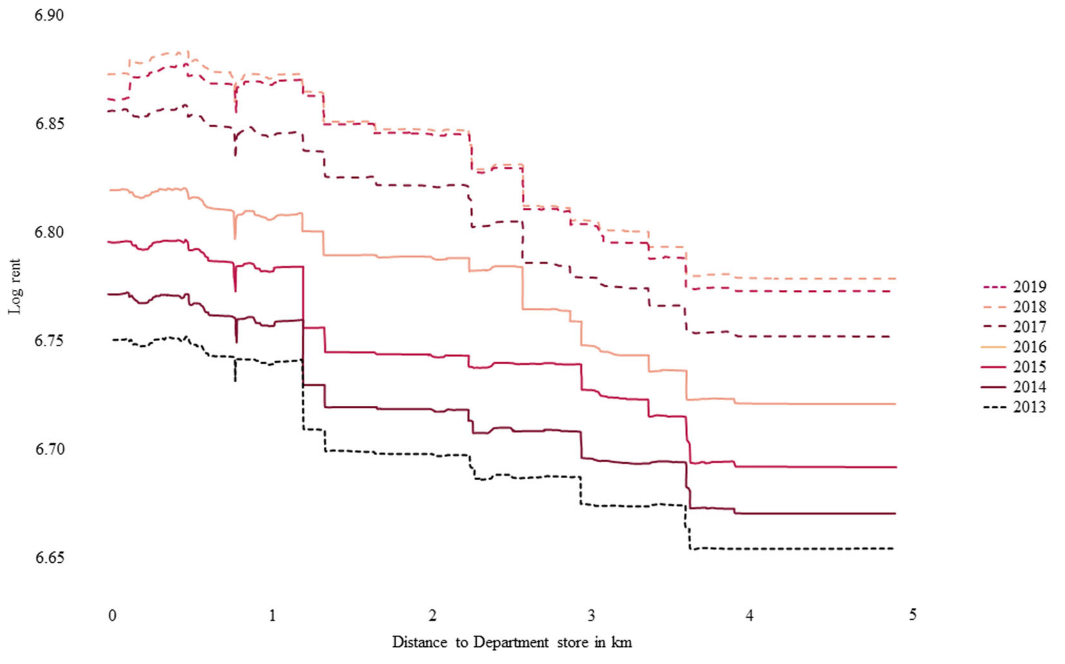


FIGURE 9 PD plot for rent and distance to department store for the years 2013 to 2019 [Color figure can be viewed at wileyonlinelibrary.com]

Note: The figure displays the partial dependence of important variables over different periods. The vertical axis denotes the feature values of the log rent level while the horizontal axis denotes the covariates feature values.

Source: Own depiction.

authors, such as Osland (2010) or Zheng et al. (2016), also find a negative relationship between property prices and distance to the city center. However, the opposite effect occurs for close proximity. We expect tenants to appreciate separation from areas that are characterized by urban disamenities such as crime, noise, and air pollution. Interestingly, apartments close to the CBD attract comparable rental values as those at a 5 km distance. A steep decrease in rent levels can be seen beyond 5 and 7.5 km.

Regarding local supply, department stores are linearly and negatively associated with rental values. The proximity to shopping facilities results in increasing rents. We do not find an equivalent distance variable in the hedonic literature, however, Dubé and Legros (2016) show a positive price effect for properties not more than 1 km away from a shopping center. Furthermore, FI ranks the presence of a built-in kitchen as important. Gröbel and Thomschke (2018) find a significant positive relationship between built-in kitchens and rents in Berlin (Germany).

Purchasing power and rent show a slightly positive correlation, indicating that rents in postcodes with strong purchasing power are higher than those in weaker ones. Neighborhoods with high purchasing power are associated with more expensive apartments and, thus, the variable is perceived as a characteristic of a good residential area. While the construction of high-rise buildings is restricted in most German cities, Frankfurt has a long-standing history of integrating tower buildings into urban planning. These do not only represent the highest price segment in the residential market of Frankfurt but have shown to be a driver of residential prices and rents in the last years.

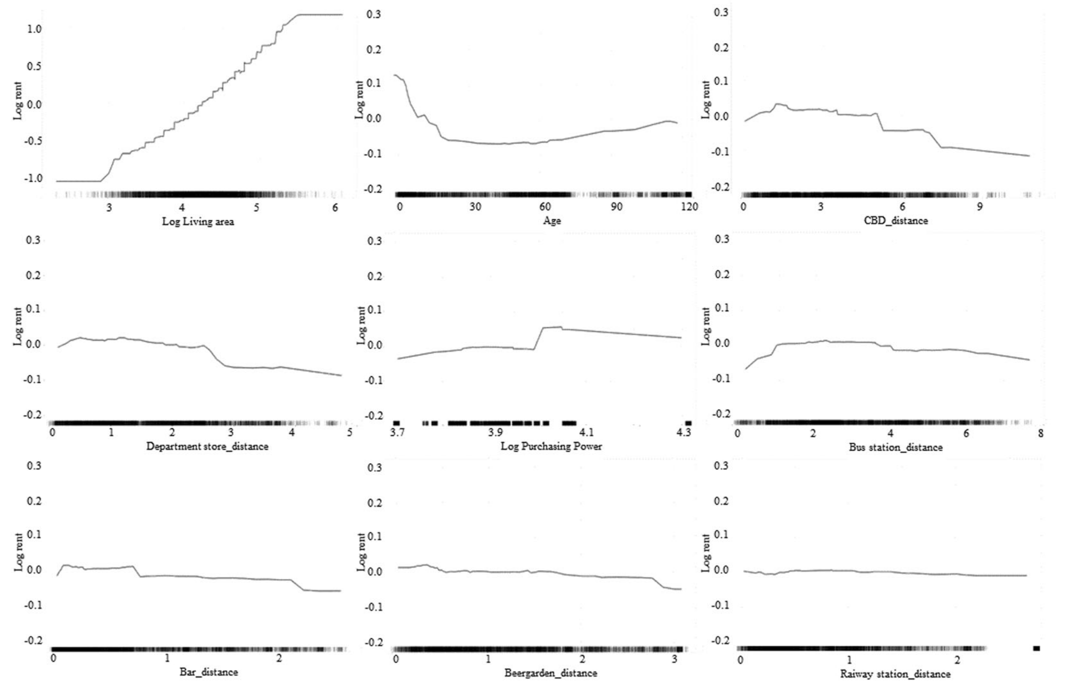


FIGURE 10 First set of ALE Plots for the year 2019
Note: The figure displays the accumulated local effects of size in square meter, age, distance to CBD, department store, Bus station, Bar, Beergarden, Railway station and Purchasing Power. The vertical axis denotes the relative feature values of log rent level while the horizontal axis represents the covariates feature values. Stacked black lines display the number of observations. *Source*: Own depiction.

Distance to hospitality and bus station affect the overall prediction the most out of all hospitality and public transport features. All three variables show a nonlinear relationship with residential rents. A bus stop in the immediate vicinity tends to have a negative effect, while a little distance away shows that the rent is increasing. This effect is then lost with even greater distance.

The plot reveals the relationship to be broadly continuous up until 3.5 km, followed by declining rental prices. Accessibility to central hubs through different means of transport seems to overlay negative effect of a larger distance. However, after 3.5 km, we find this effect to become visible and apartments that are poorly located in terms of transport face discounts for low accessibility. The proximity to hospitality is mostly positive. We find distance to a bar or pub to be positively associated with rental values up to approximately 700 m. While a bar in very close proximity would result in lower rents, the access to hospitality leads to an increase in rental values only from a certain distance. We expect tenants to face a trade-off between accessibility and negative externalities such as noise. In contrast, the presence of a train station has no significance for rental potential, at least based on visual analysis.

Figure 6 provides another nine features with their PD plots. Generally, it should be noted that the effects of the features tend to attenuate with distance. This finding is consistent with the FI of the respective properties. While there is a weakly positive link between the distance to a car wash or to traffic facilities and the apartment rent, other variables such as the distance to a park or supermarket, show a neutral linear relationship. Consequently, we refrain from further verbal analysis of this graph and focus on the observation.

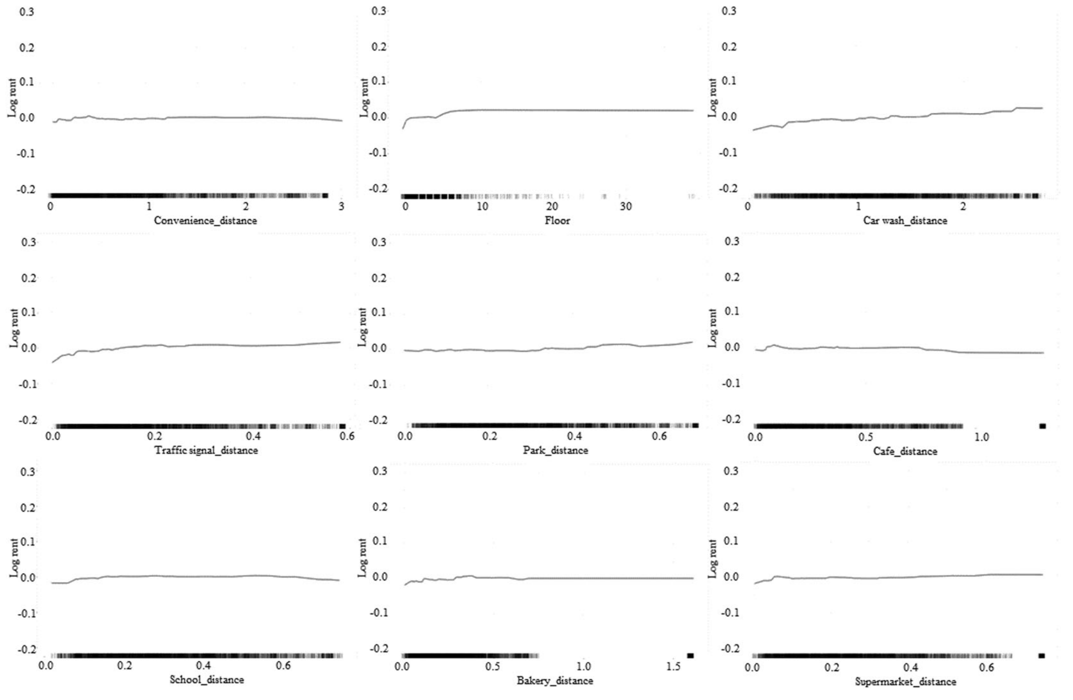


FIGURE 11 Second set of ALEs of certain features in 2019
Note: The figure displays the accumulated local effects of the variables Distance to convenience store, floor, distance to car wash, traffic signal, park, café, school, bakery and supermarket. The vertical axis denotes the relative feature values of log rent level while the horizontal axis represents the covariates feature values. Stacked black lines display the number of observations.
Source: Own depiction.

Regarding the binary variables displayed in Figure 7, we find the appearance of a built-in kitchen to affect the rental level most. Furthermore, having a parking sport, an elevator as well as a balcony increase the rent. In contrast, an apartment with a bathtub face rental discounts.

Since the previous metric and binary PD analyses are limited to 2019 only, two variables are considered over time below. Adding a temporal dimension to our analysis by displaying feature effects on a yearly basis enables us in a last step to illustrate temporal dynamics of the effects of hedonic characteristics. We demonstrate the latter by analyzing the distance to the CBD and the distance to a department store (both previously presented in Figure 5). Moreover, we rescale the plot regarding the y-axis to better visualize the individual effects.

At first, Figure 8 shows a negative relationship between rents and the distance to CBD across time. A continuous upwards shift for all feature values indicates increasing rent levels during the observed period. Only the graph of the year 2019 behaves differently since it moves below 2018 for closer proximity and analogous from 5 km distance onwards. This development could be attributed to a declining preference for downtown locations in combination with overall stable rent levels in recent years. Although the trajectory of all lines is similar, we find some differences. First, a drop in rental prices at a distance of 5 km is less pronounced for 2017, 2018, and 2019 than for previous years. Second, another major decline can be recognized at 7 km for 2013 to 2016. In the following years 2017 to 2019, however, this is only noticeable at approximately 7.5 km but the downturn is considerably stronger. Both changes indicate that residential locations in medium

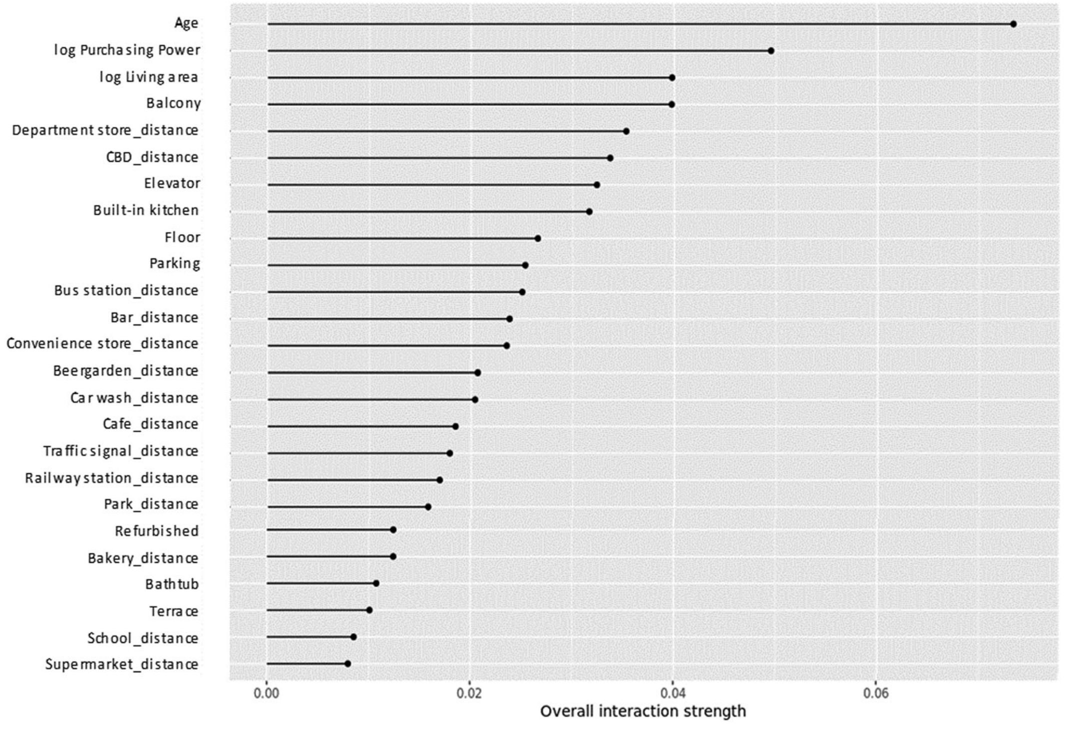


FIGURE 12 Interaction strength of the individual features
Note: The figure displays the interaction strength on the x-Axis of the individual features on the y-Axis.
Source: Own depiction.

distance to the center (5 to 7.5 km) experienced stronger rent increases compared to central as well as periphery location.

In Figure 9, a negative relationship between rents and distance to a department store is displayed. Apart from a general rental growth pattern over the years, we observe a drop in achievable rents at approximately 1.2 km from a major shopping facility, with the years 2013, 2014, and 2015 experiencing a stronger drop in the rental gradient with several other steeper declines at further distances. The findings indicate that locations between 1.2 and 2.8 km appreciated more rapidly relative to locations both in closer proximity as well as far away from department stores. Appendix C provides additional and centered PD plots for the feature Distance to department store. Centered PD plots aid and underpin the interpretation of the differences in PDs throughout the years.

Generally, the interpretation of PD plots can be biased when features are highly correlated, as their computation is based on unrealistic combinations of feature values for which the average prediction is calculated. To overcome these issues, we use ALE plots as an alternative to visualizes feature effects.

Using conditional distributions and the difference of predictions instead of the average reveals the pure effect that is solely based on the specific variable of interest. In our case, the ALE analysis not only helps to show a certain robustness of the PD results, but also offers another possibility to investigate feature effects.

In general, the results of the PD and ALE plots are largely the same. However, minor differences occur. For example, is the increase in rental values in Figure 10 for small apartments higher when

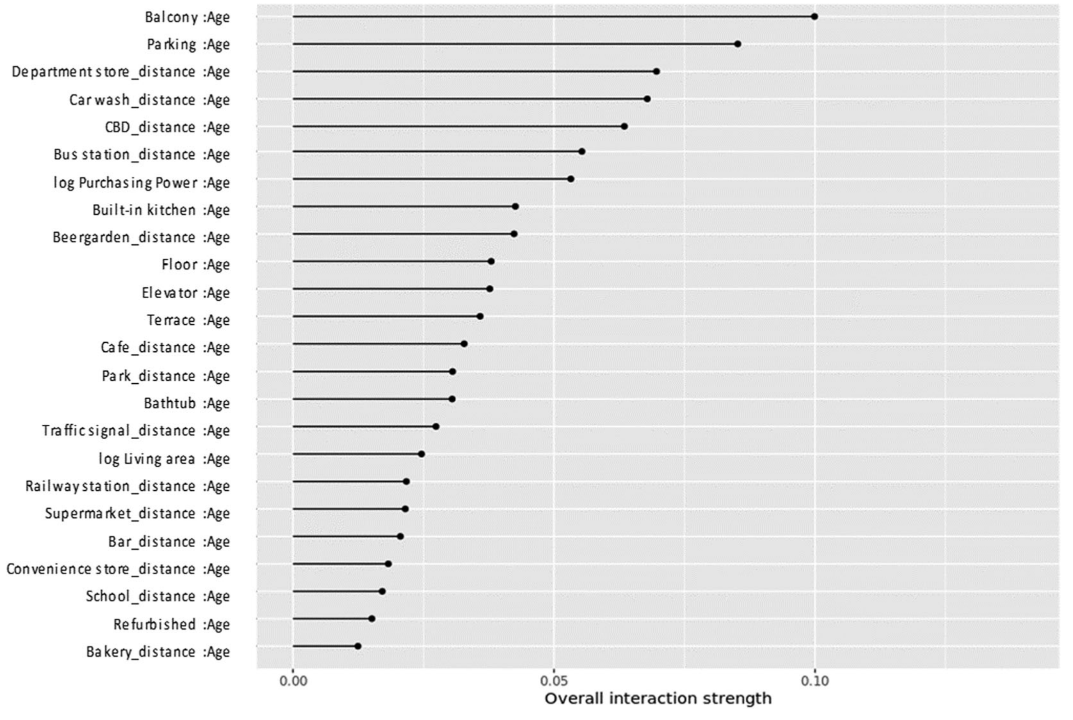


FIGURE 13 Interaction strength of the variable age
Note: The figure displays the interaction strength of the variable age on the x-Axis with remaining features on the y-Axis.
Source: Own depiction.

controlling for feature correlation in ALE plots. In contrast, the increase is less pronounced for old buildings. Furthermore, we find close proximity as well as locations far away from the CBD to face higher rental discounts. Further variables in Figure 11 do not show remarkable variations.

Although the differences are minimal in our analysis, ALE plots are an important extension to correctly visualize and interpret feature effects. Ultimately, the feature effects technique yields greater transparency of how the different inputs contribute to the final estimation of the ML model. By visualizing the individual relations between the variables and the rent to be estimated, this method demonstrates which (economic) rational the algorithm has learned from the data and accordingly integrated into its internal calculations.

5.5 | Feature interactions of the hedonic characteristics

Tree-based models capture the interaction of features due to their nested if-else structure and build "smart" bundles of attributes. Therefore, it is possible to show whether and how strongly a single feature interacts with all other features in the model. An overall interactions strength measures to which extend the interaction of one feature with all other features in the model explains the variance of the ML model f. The amount of variance explained (interaction strength) ranges from 0 (no interaction) to 1 (fully explained by interactions). The individual variable interaction is shown in Figure 12. In general, the interaction strength for all features is low, with none of the variables exceeding 8% of the variance to be explained by the interaction per feature. The results

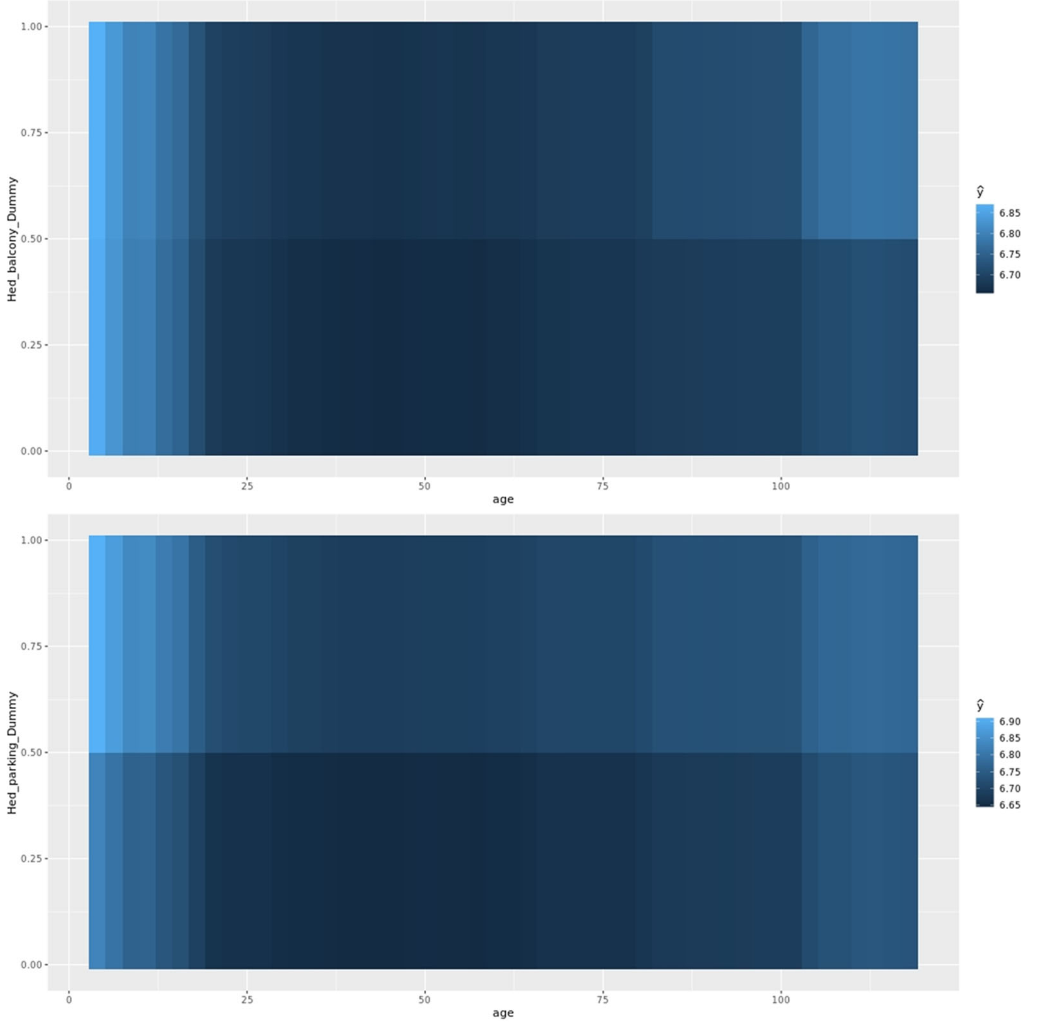


FIGURE 14 Two-dimensional PD Plot of the variables age and balcony/parking
[Color figure can be viewed at wileyonlinelibrary.com]

Note: The figure displays the 2-dimensional partial dependence of the variables balcony and parking in combination with age. The y-axis denotes the feature value of the variables balcony respective parking. The x-axis denotes the feature values of the variable age. While darker areas indicate low estimated rental values, lighter ones mark high levels.
Source: Own depiction.

indicate that age, purchasing power per household, as well as the Existence of a balcony interacts most with other features.

Figure 13 illustrates the two-way interaction of the characteristic age with all other features. Interestingly, balcony as well as the presence of a parking spot show high interaction with the building age. These findings reflect historical city development of Frankfurt as well as planning law regulations. While historical buildings in the city center where mostly built without balcony and parking lots, new buildings face requirements on the minimum number of parking lots. Since this metric solely focuses on the strength without showing the interaction between the two features, we conduct two-dimensional PD plots to analyze these two relationships in more detail.

Figure 14 shows the two-dimensional PD plots for age and both balcony (former) and parking (latter graphic). The plots are based on the known methodological framework of PD plots, with the subset of features XS now comprising two features. The plots illustrate how the rent level (light blue indicating higher rental values) depends on the age as well as the presence of a balcony and a parking slot, respectively. Period properties without a balcony are discounted while on-site parking is found to enhance the rents of new builds.

The analysis of feature interactions does not only allow us to highlight how ML models incorporate interactions between features within their tree-based structure but also yields information on the importance of different interacting variable. This could be the basis for different terms to be included in further models such as parametric or semiparametric terms.

6 | CONCLUSIONS

This article set out to demonstrate how ML -based decision making in hedonic modeling can be made more transparent to rearchers and practitioners. We visualize and investigate the relationship between residential rents and a set of hedonic variables using an ML model. Based on a sample of 52k apartments in Frankfurt am Main, Germany, we apply parametric and semiparametric models as well as the XGB algorithm for rental prediction. In a next step, model-agnostic IML methods are used to examine FI, feature effects, and interactions.

FI reveals and ranks the individual importance of variables for rental prediction. In our empirical analysis, size and age have the greatest influence on rents. Considering correlated features, hierarchical clustering reveals that distance variables must be viewed with caution when using interpretive methods. Feature effects illustrate the relationships between rent and the influencing variables using PD and ALE. For example, the relationship of size to rent is linear whereas age and other variables show nonlinear dependencies. In addition, the inspection of PD plots on a yearly basis shows that relationships can also change over time. We also examine how individual features interact with each other in the model. For example, private parking and a balcony in conjunction with historic building status affects the rent above and beyond the sum of the unconditional individual effects.

Our findings demonstrate that interpretation methods can reveal the mechanisms behind the ML estimates by highlighting the relationship that the algorithm detects in the underlying data. Peeking inside the black box with these methods enables researchers to understand how a ML model arrived at its prediction and will help to gain new insights, ease practical applications, and enhance reliability in algorithmic decisions.

The insights gleaned from these methods are relevant not only for research but also for private and public sector practice. Real estate professionals commonly use ML in their decision-making process (RICS, 2017). Model-agnostic methods may improve the interpretability of AI-based results for investment purposes compared to more rigid mutivariate regression specfiations or black box forecasting models. Whereas the advantages of these methods have already been discussed in detail, difficulties and limitations must also be pointed out. First, ML-based methods, such as XGB as well as their interpretation with IML, face challenges in terms of computing power due to the complex algorithm structures. Second, it should be noted that data availability is of course essential for hedonic models. Even with ML-based models, an omitted variable bias can drastically reduce the informative value and thus the applicability.

IML is a rapidly evolving field as new methods and applications are being continuously proposed in a variety of applications. Although this research area has by now achieved a moderate degree of stability (Molnar et al., 2020), it is still in its infancy and still faces several challenges to overcome. To be of practical use in real estate decision making, more research is needed to improve the reliability and validity of algorithmic decision making. We expect IML methods to be a valuable addition to established quantitative appraisal and research methods both because they contribute to the transparency of ML models and provide insights into potentially unknown relationships in real estate hedonic modeling.

ACKNOWLEDGMENTS

The authors thank PATRIZIA AG for contributing to this research. All statements of opinion reflect the current estimations of the authors and do not necessarily reflect the opinion of PATRIZIA AG or its associated companies. Furthermore, the authors would like to thank all parties involved especially the researchers and colleagues at the conferences as well as the editor and referees who provided many helpful comments. Moreover, Franz Fuerst wishes to acknowledge support of his research from the Cambridge University Land Society.

ORCID

Felix Lorenz https://orcid.org/0000-0002-9795-7700
Jonas Willwersch https://orcid.org/0000-0003-3102-9044
Marcelo Cajias https://orcid.org/0000-0002-0777-7459
Franz Fuerst https://orcid.org/0000-0003-2780-938X

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160.

Ahmad, M. A., Eckert, C., & Teredesai, A. (2018). Interpretable machine learning in healthcare. *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 559–560.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Springer Science & Business Media.

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: an application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, *39*(2), 1772–1778.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(4), 1059–1086.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115.

Below, S., Beracha, E., & Skiba, H. (2015). Land erosion and coastal home values. *Journal of Real Estate Research*, *37*(4), 499–536.

Berk, R. A., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior: a comparative assessment. *Criminology & Public Policy*, *12*, 513–544.

Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, *9*(1), 7–27.

Bontemps, C., Simioni, M., & Surry, Y. (2008). Semiparametric hedonic price models: assessing the effects of agricultural nonpoint source pollution. *Journal of Applied Econometrics*, *23*(6), 825–842.

Breiman, L. (2001a). Random forests. *Machine Learning*, *45*(1), 5–32.

Breiman, L. (2001b). Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231.

Brooks, T. J., Humphreys, B. R., & Nowak, A. (2020). Strip clubs, "secondary effects" and residential property prices. *Real Estate Economics*, *48*(3), 850–885.

Brunauer, W., Lang, S., & Umlauf, N. (2013). Modelling house prices using multilevel structured additive regression. *Statistical Modelling*, *13*(2), 95–123.

Cajias, M., & Freudenreich, P. (2018). Exploring the determinants of liquidity with big data–market heterogeneity in German markets. *Journal of Property Investment & Finance*, *36*(1), 3–18.

Cajias, M., Willwersch, J., Lorenz, F., & Schaefers, W. (2021). Rental pricing of residential market and portfolio data–a hedonic machine learning approach. *Real Estate Finance*, *38*(1), 1–17.

Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, *22*(3), 453–474.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: a survey on methods and metrics. *Electronics*, *8*(8), 832.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 785–794.

Chernobai, E., Reibel, M., & Carney, M. (2011). Nonlinear spatial and temporal effects of highway construction on house prices. *Journal of Real Estate Finance and Economics*, *42*(3), 348–370.

Chin, S., Kahn, M. E., & Moon, H. R. (2020). Estimating the gains from new rail transit investment: a machine learning tree approach. *Real Estate Economics*, *48*(3), 886–914.

Chrostek, K., & Kopczewska, K. (2013). Spatial prediction models for real estate market analysis. *Ekonomia*, *35*(0).

Conway, D., Li, C. Q., Wolch, J., Kahle, C., & Jerrett, M. (2010). A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values. *Journal of Real Estate Finance and Economics*, *41*(2), 150–169.

Court, A. T. (1939). Hedonic price indexes with automotive examples. In *The dynamics of automobile demand* (pp. 99–117). New York: General Motors Corporation.

Des Rosiers, F., Dubé, J., & Thériault, M. (2011). Do peer effects shape property values? *Journal of Property Investment & Finance*, *29*(4/5), 510–528.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *Working Paper. ArXiv:1702.08608*.

Dubé, J., & Legros, D. (2016). A spatiotemporal solution for the simultaneous sale price and time-on-the-market problem. *Real Estate Economics*, *44*(4), 846–877.

Dubin, R. A. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *Review of Economics and Statistics*, *70*, 466–474.

Dumm, R. E., Sirmans, G. S., & Smersh, G. T. (2016). Price variation in waterfront properties over the economic cycle. *Journal of Real Estate Research*, *38*(1), 1–26.

Dumm, R. E., Sirmans, G. S., & Smersh, G. T. (2018). Sinkholes and residential property prices: presence, proximity, and density. *Journal of Real Estate Research*, *40*(1), 41–68.

Fernández-Avilés, G., Minguez, R., & Montero, J.-M. (2012). Geostatistical air pollution indexes in spatial hedonic models: the case of Madrid, Spain. *Journal of Real Estate Research*, *34*(2), 243–274.

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1–81.

Freybote, J., Fang, Y., & Gebhardt, M. (2017). The impact of temporary uses on property prices: the example of food trucks. *Journal of Property Research*, *34*(1), 19–35.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232.

Goodwin, K. R., La Roche, C. R., & Waller, B. D. (2020). Restrictions versus amenities: the differential impact of home owners associations on property marketability. *Journal of Property Research*, *37*(3), 238–253.

Greenwell, B. M. (2017). pdp: an R package for constructing partial dependence plots. *R Journal*, *9*(1), 421–436.

Gröbel, S. (2019). Analysis of spatial variance clustering in the hedonic modeling of housing prices. *Journal of Property Research*, *36*(1), 1–26.

Gröbel, S., & Thomschke, L. (2018). Hedonic pricing and the spatial structure of housing data—an application to Berlin. *Journal of Property Research*, *35*(3), 185–208.

Haider, M., & Miller, E. J. (2000). Effects of transportation infrastructure and location on residential real estate values: application of spatial autoregressive techniques. *Transportation Research Record*, *1722*(1), 1–8.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Ho, W. K., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, *38*(1), 48–70.

Hoen, B., & Atkinson-Palombo, C. (2016). Wind turbines, amenities and disamenitites: a study of home value impacts in densely populated Massachusetts. *Journal of Real Estate Research*, *38*(4), 473–504.

Hoen, B., Brown, J. P., Jackson, T., Thayer, M. A., Wiser, R., & Cappers, P. (2015). Spatial hedonic analysis of the effects of US wind energy facilities on surrounding property values. *Journal of Real Estate Finance and Economics*, *51*(1), 22–51.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (6th ed.). Springer.

Jauregui, A., Allen, M. T., & Weeks, H. S. (2019). A spatial analysis of the impact of float distance on the values of canal-front houses. *Journal of Real Estate Research*, *41*(2), 285–318.

Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big data in real estate? From manual appraisal to automated valuation. *Journal of Portfolio Management*, *43*(6), 202–211.

Kontrimas, V., & Verikas, A. (2011). The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing*, *11*(1), 443–448.

Lam, K. C., Yu, C. Y., & Lam, C. K. (2009). Support vector machine and entropy based decision support system for property valuation. *Journal of Property Research*, *26*(3), 213–233.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132–157.

Lechner, M., Hasani, R., Amini, A., Henzinger, T. A., Rus, D., & Grosu, R. (2020). Neural circuit policies enabling auditable autonomy. *Nature Machine Intelligence*, *2*(10), 642–652.

LeSage, J. (1999). The theory and practice of spatial econometrics. *University of Toledo. Toledo, Ohio*, *28*(11), 1–39.

Li, T. (2020). The value of access to rail transit in a congested city: evidence from housing prices in Beijing. *Real Estate Economics*, *48*(2), 556–598.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: a review of machine learning interpretability methods. *Entropy*, *23*(1), 18.

Malpezzi, S. (2002). Hedonic pricing models: a selective and applied review. In T. O'Sullivan & K. Gibb (Eds.), *Housing economics and public policy* (pp. 67–89). Blackwell Science Ltd.

Mason, C., & Quigley, J. M. (1996). Non-parametric hedonic housing prices. *Housing Studies*, *11*(3), 373–385.

Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*, *12*(1), 134–150.

McCluskey, W. J., McCord, M., Davis, P. T., Haran, M., & McIlhatton, D. (2013). Prediction accuracy in mass appraisal: a comparison of modern approaches. *Journal of Property Research*, *30*(4), 239–265.

Molnar, C. (2020). *Interpretable machine learning*. A guide for making Black *box models expl*ainable.

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges. *Working Paper ArXiv:2010.09337*.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, *32*(3), 289–320.

Pace, R. K. (1998). Appraisal using generalized additive models. *Journal of Real Estate Research*, *15*(1), 77–99.

Pace, R. K., & Hayunga, D. (2020). Examining the information content of residuals from hedonic and spatial models using trees and forests. *Journal of Real Estate Finance and Economics*, *60*(1-2), 170–180.

Pérez-Rave, J. I., Correa-Morales, J. C., & González-Echavarría, F. (2019). A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes. *Journal of Property Research*, *36*(1), 59–96.

R Core Team. (2020). *R: A language and environment for statistical computing*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-Agnostic Interpretability of Machine Learning. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*.

RICS. (2017). *The Future of Valuations*. Royal Institute of Chartered Surveyor.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55.

Rouwendal, J., Levkovich, O., & van Marwijk, R. (2017). Estimating the value of proximity to water, when ceteris really is paribus. *Real Estate Economics*, *45*(4), 829–860.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, *577*(7792), 706–710.

Seo, W. (2018). Does neighborhood condition create a discount effect on house list prices? Evidence from physical disorder. *Journal of Real Estate Research*, *40*(1), 69–88.

Sheppard, S. (1999). Hedonic analysis of housing markets. In P. Cheshire & E. S. Mills (Eds.), *Applied urban economics: Vol. 3. Handbook of regional and urban economics* (pp. 1595–1635). Elsevier.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The composition of hedonic pricing models. *Journal of Real Estate Literature*, *13*(1), 1–44.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222.

Stamou, M., Mimis, A., & Rovolis, A. (2017). House price determinants in Athens: a spatial econometric approach. *Journal of Property Research*, *34*(4), 269–284.

Theisen, T., & Emblem, A. W. (2018). House prices and proximity to kindergarten–costs of distance and external effects? *Journal of Property Research*, *35*(4), 321–343.

van Wezel, M., Kagie, M. M., & Potharst, R. R. (2005). Boosting the accuracy of hedonic pricing models. Econometric Institute, Erasmus University Rotterdam. http://hdl.handle.net/1765/7145

Wyman, D., & Mothorpe, C. (2018). The pricing of power lines: a geospatial approach to measuring residential property values. *Journal of Real Estate Research*, *40*(1), 121–154.

Yao, Y., Zhang, J., Hong, Y., Liang, H., & He, J. (2018). Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. *Transactions in GIS*, *22*(2), 561–581.

Yoo, S., Im, J., & Wagner, J. E. (2012). Variable selection for hedonic model using machine learning approaches: a case study in Onondaga County, NY. *Landscape and Urban Planning*, *107*(3), 293–306.

Zhao, Q., & Hastie, T. (2021). Causal interpretations of Black-Box Models. *Journal of Business & Economic Statistics*, *39*(1), 272–281.

Zheng, S., Hu, W., & Wang, R. (2016). How much is a good school worth in Beijing? Identifying price premium with paired resale and rental data. *Journal of Real Estate Finance and Economics*, *53*(2), 184–199.

Zurada, J., Levitan, A., & Guan, J. (2011). A comparison of regression and artificial intelligence methods in a mass appraisal context. *Journal of Real Estate Research*, *33*(3), 349–387.

How to cite this article: Lorenz, F., Willwersch, J., Cajias, M., & Fuerst, F. (2023). Interpretable machine learning for real estate market analysis. *Real Estate Economics*, *51,* 1178–1208. https://doi.org/10.1111/1540-6229.12397

APPENDIX A: METHODOLOGY ON PARAMETRIC AND SEMIPARAMETRIC APPROACHES

We apply different hedonic methods that have been used in the extant literature. First, we deploy a hedonic OLS modeling approach to estimate the effects of property characteristics on rental prices. Linear hedonic regression represents the standard approach in modeling real estate prices and rents and is frequently used in housing studies (Mayer et al., 2019). The hedonic regression describes the rent Y as the sum of the predicted values of its characteristics X_j:

$$Y = \beta_0 + \sum_{j=1}^{J} X_j \beta_j + \varepsilon. \tag{8}$$

In accordance with the real estate literature, a semi-log functional form with log-transformation of the dependent variable is conducted. Property characteristics include structural, socioeconomic neighborhood, and locational features. Proximity variables account for the spatial distance to public amenities and transport. Further spatial effects are modeled via spatial expansion by incorporating the coordinates in terms of longitude and latitude (Bitter et al., 2007; Chrostek & Kopczewska, 2013; Pace & Hayunga, 2020). Furthermore, temporal dummies are included for the specific month and year.

Many authors argue that property prices and rents may contain two key figures, namely spatial autocorrelation and spatial heterogeneity that can require the spatial extension of hedonic models (LeSage, 1999). Since the occurrence of spatial effects can lead to misspecifications and biased results in the OLS framework (Anselin, 1988), we additionally apply a SAR with the following functional form:

$$Y = X\beta + \rho WY + \varepsilon. \tag{9}$$

ρWY denotes a spatial lag of the target variable Y with W being the spatial weight matrix that specifies the spatial structure, and ρ representing the spatial lag parameter.

However, linear models are subject to various restrictions due to their functional parametric form that can yield to misspecifications (Mason & Quigley, 1996; Pace, 1998). Because relationships in housing markets appear often to be nonlinear, hedonic modeling can require the incorporation of more flexible functional forms to account for nonlinearity (Bontemps et al., 2008; Brunauer et al., 2013). Hence, a semiparametric generalized additive model (GAM) is further considered.

$$Y = \beta_o + \sum_{j=1}^{J} X_j \beta_j + \sum_{p=1}^{P} f_p \left(X_p\right) + \varepsilon, \tag{10}$$

GAM relaxes the linearity assumption by replacing the parametric linear relationship with non-parametric smoothers (e.g., splines, near neighbor and kernel smoothers). The linear equation is expanded by p smooth functions f_p to identify latent nonlinear effects.

APPENDIX B: XGB AND PARAMETER TUNING

ML models can identify complex pattern due to their high flexibility. By learning from the underlying data and iteratively improving the final model, ML methods yield remarkable predictive results. To avoid that the model fits the observations within the sample perfectly but fails to predict unseen data (out-of-sample), the sample is divided into test and training set, with the model learning from the training set and being evaluated using the test set.

The performance of the ML model is dependent on the right choice of hyperparameter. The applied XGB algorithm is based on the following parameters: Max depth is the maximum depth of a tree that is allowed. A deeper tree allows the model to be more complex. γ is the minimum split loss and defines minimum loss reduction required to a make a split. XGB involves a so-called regularized objective $\mathcal{L}(\phi)$ that penalizes complex models. Both parameters η and λ control for the complexity of the regulation. A lower learning rate η shrinks the learning process and makes the boosting process more conservative. α and λ are both regularization terms (L1 and L2 regularization) to counteract over-fitting. We pre-define the search space for every parameter. The hyperparameter are tuned using random search. K-fold cross validation is applied to evaluate the tuning process and find the optimal combinations of parameters. The number of folds is chosen to be five.

Results from ML models can be very sensitive to the final set of hyperparameter. Although this is less pronounced for XGB than for other models, we want to ensure that hyperparameter tuning does not affect the interpretability of the results. Since the increase in model performance highly decelerates after approximately 40 iterations, it is important to verify that different sets of hyperparameter do not results in different interpretations. To show the robustness of the results, we randomly choose a second set of parameters in addition to the final parameter choice. Therefore, we pick the iteration 87 in Figure 2. For this iteration, the maximum depth is higher, with 42. Furthermore, the model is trained with higher regularization parameters η, α, and λ. Figure 15 shows the PD plot for the feature age for both sets of hyperparameters. We do not find the changes in hyperparameters to affect the interpretability of the results. Only, if we visualized both plots on a higher scale than used in Figure 5, we find marginal differences in the final outcome.

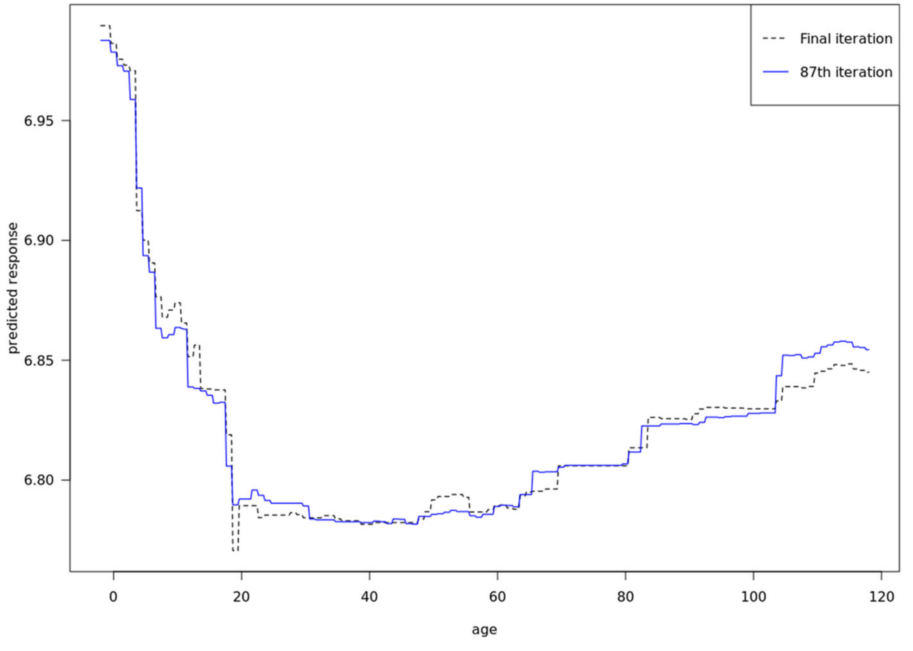
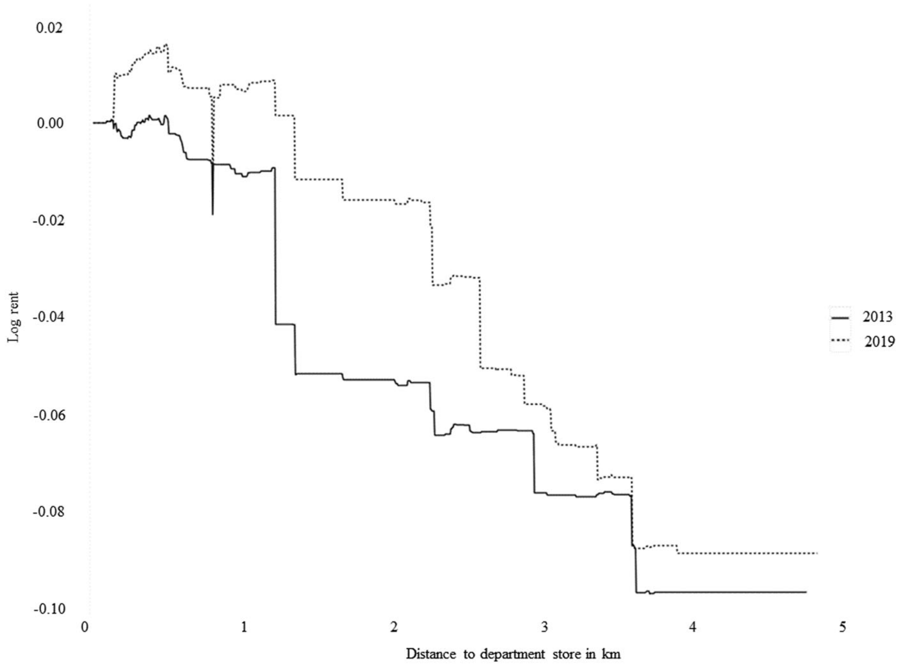


FIGURE 15 PD Plot of the variables age for two sets of hyperparameters
[Color figure can be viewed at wileyonlinelibrary.com]

Note: The figure displays the partial dependence of the variable Age for two sets of hyperparameters. The vertical axis denotes the feature values of log rent level while the horizontal axis represents the covariates feature values. The stacked line represents the 87th iteration within the tuning process. The solid line indicates the finale set of hyperparameters after 100 iterations.
Source: Own depiction.

APPENDIX C: CENTERED PD PLOT FOR DISTANCE TO DEPARTMENT STORE



Note: The figure displays the partial dependence centered at lowest feature value. The vertical axis denotes the feature values of the log rent level while the horizontal axis denotes the covariates feature values.
Source: Own depiction.