AGILE: GIScience Series, 7, 15, 2026. https://doi.org/10.5194/agile-giss-7-15-2026
Proceedings of the 29th AGILE Conference on Geographic Information Science, 16–19 June 2026.
Eds.: Evelyn Uuemaa, Alexander Kmoch, Holger Virro, and Simon Scheider
This contribution underwent peer review based on a full paper submission.

The value of where – quantifying location values with tabular transformer models

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. Location is a key feature of urban spaces. It influences housing and land markets, social segregation, and decisions in urban planning and administration. Hedonic price models and, increasingly, machine learning (ML) methods are traditionally used for quantitative location valuation. At the same time, transformer architectures and their more recent adaptations for tabular data mark a methodological milestone. However, to date, tabular transformers have not been widely adopted in spatial analyses and, in particular, have not been systematically evaluated for location value assessments. This study addresses this gap by comparing classic ML methods with two tabular transformer approaches: TabNet and Mitra. In a case study, rents are used as the target variable, with property-specific characteristics being normalised using spatial regressions. Location-describing features include accessibility to POIs and areal variables such as imperviousness, noise, and population density. While a stacked ensemble of classical ML models achieved the lowest RMSE, TabNet and Mitra performed within a narrow margin, demonstrating that they can be used to map spatial patterns.

Submission Type. Case study, model comparison

BoK Concepts. [GC3] Artificial intelligence (AI) in EO and GI, [GC1] Geocomputation and complex systems

Keywords. AI in GIS, Spatial Data Science, Urban Analytics, Tabular Transformer

1 Introduction

The monetary value of ‘location’ is a key means of managing and describing urban processes. Location has a significant influence on how the urban system functions because it expresses accessibility, environmental pollution, and proximity to urban amenities and their uneven spatial distribution, thereby influencing planning processes that affect mobility, location decisions, social participation, and land use dynamics. From a hedonic perspective, location is not a single characteristic, but a bundle of implicit attributes (accessibility, environmental quality, social environment) whose willingness to pay for is ‘revealed’ in market prices (Rosen, 1974). In recent urban research, this idea has been used, amongst other things, to examine accessibility as a core indicator of urban quality of life, sustainability and participation, for example in discussions about the ‘15-minute city’, in which the accessibility of everyday destinations close to home serves as a planning model and heterogeneities are discussed as a question of social justice (Moreno et al., 2021; Pereira et al., 2017). This logic is not only theoretical, but also shapes concrete planning and financing issues in urban systems. For example, it has been shown that improved public transport connections are capitalised by translating into higher prices near stations (Debrezion et al., 2007; Gibbons & Machin, 2005). Similar price effects can be seen with improvements in urban green space coverage (Panduro & Veie, 2013) and reductions in noise pollution (Day et al., 2007).

In many European countries, the consideration of location and its calculation are anchored in law. The Spanish Housing Act, for example, introduced a state reference system for rental prices in which location plays a central role (Jefatura del Estado, 2023). In Austria's partially regulated tenancy law, rents are subject to location surcharges and discounts (Mietrechtsgesetz, 1981). In Germany, the national ImmoWertV cites location as a value-relevant feature (ImmoWertV, 2021), and for qualified rent indices (MsV, § 19), the BGB expressly requires the application of recognised scientific principles (BGB, § 558d). Against this background, it is logical to evaluate the modelling of location values using modern tabular transformer approaches compared to established regression and machine learning (ML) methods.

2 Related works

2.1 Regression

In real estate economics, the hedonic price model is standard: the price is modelled as a function of the ‘bundling’ of characteristics, e.g., living space, condition, accessibility, environment (Rosen, 1974; Can, 1992). A key advantage of this approach is the interpretability of the models, but at the expense of predictive power relative to more modern methods (Breiman, 2001a). In addition, these models generally neglect the spatial nature of the underlying data.

This is addressed by spatial autoregressive regression (SAR). These models capture spatial interaction and allow for more realistic modelling of spatial dependencies in rental and purchase data (Basu & Thibodeau, 1998; Anselin, 1988).

In this study, SAR is used to normalise property characteristics and distil location effects. Property characteristics are ‘calculated out’ so that the structure-adjusted prices represent the location value more realistically.

2.2 Traditional machine learning methods

In addition to traditional regression approaches, ML methods are increasingly used in rent and property valuation as well as location value calculations, particularly because they can model non-linear relationships and complex feature interactions (e.g., Park & Bae, 2015; Adetunji et al., 2022). Support vector machines (SVMs) enable this via kernel methods in high-dimensional feature spaces (Cortes & Vapnik, 1995). Random forests combine many decision trees trained on bootstrap samples (bagging) with random feature selection for splits (Breiman, 2001b) and are considered a robust non-linear standard approach in this field. Gradient boosting methods sequentially build an additive model from weak learners, with each step reducing the errors of the previous stage (Friedman, 2001). Neural networks approximate complex functions via layered transformations and are used in hedonic applications as a flexible alternative to regression (e.g., Selim, 2009).

In more recent work, these methods are also often combined as ‘stacks’: a meta-model learns from out-of-fold predictions of heterogeneous base models to compensate for systematic errors in individual models (Wolpert, 1992).

Overall, such ML approaches often achieve higher predictive performance than classical regressions, but are typically less interpretable. Post-hoc explanation methods such as LIME and SHAP can partially close this gap by quantifying local or additive feature contributions (Ribeiro et al., 2016; Lundberg & Lee, 2017).

2.3 Transformer models

Transformer models are now a key model class for learning patterns in complex, high-dimensional feature spaces and can therefore also be relevant to data-driven location assessments and spatial analyses in general.

In principle, transformers are based on self-attention (Vaswani et al., 2017). Query, key, and value vectors are calculated for each input token (e.g., word fragment, image patch, or table column). Attention then weights which other tokens are important for the current representation (e.g., a location in space). Multi-head attention combines several ‘views’ on these dependencies. In addition, so-called residual connections and normalisation stabilise the training (Vaswani et al., 2017). The strengths of this approach lie in (i) the flexible modelling of global interactions, (ii) good scalability with many data points and parameters, and (iii) pre-training on large amounts of data and subsequent transfer to specific areas of application.

After this approach was presented in 2017 (Vaswani et al., 2017), BERT demonstrated the benefits of masked pre-training for representation learning in 2018/2019 (Devlin et al., 2019), and GPT-3 demonstrated the scaling effects of large generative models in 2020 (Brown et al., 2020).

According to a scope search on pertinent literature databases, despite their rapid adoption in other fields, transformers have rarely been used for location estimation to date (but see Moghimi et al., 2023; Zhao et al., 2023), partly due to (i) data protection concerns when using cloud/API services from proprietary providers, especially when analysing microdata, and (ii) because transformers were originally optimised for text generation and not for predicting spatially continuous variables. However, (i) locally operated open-weight models can be used, and (ii) tabular transformers directly address tabular structures as they exist in attributed spatial datasets (e.g., TabTransformer, FT-Transformer, SAINT) (Huang et al., 2020; Gorishniy et al., 2021; Somepalli et al., 2021).

This study considers TabNet, an approach classified as a tabular transformer, despite some differences from the classic architecture. The basic idea is similar to gradient boosting, but instead of building on tree-based models, TabNet uses neural networks. It combines sequential feature selection across multiple decision steps (similar to gradient boosting) with sparsity-regularised attention masks (similar to transformers) to utilise distinct feature subspaces at each step (Arik & Pfister, 2021).

Advantages of this approach include often strong performance without feature engineering and built-in, limited feature selection signals that can be used for interpretation. However, TabNet is sensitive to hyperparameters, which can lead to unstable training.

The second model considered in this study is Mitra, a newer approach than TabNet. It is a pre-trained, transformer-based tabular foundation model based on In-Context Learning (ICL) (Zhang et al., 2025). The model was pre-trained with considerable computational effort on synthetic tabular tasks derived from a curated mixture of data-generating priors with ~45 million datasets. These priors include Structural Causal Models (SCMs) and Tree-Based Priors (TBP), enabling the model to map different dependency structures and decision rules based on thresholds. For predictions, the model receives a support set (features and labels) and a query set (features only), and generates predictions conditioned on the in-context examples during a forward pass. In addition, this foundation model can be fine-tuned with application-specific data by further optimising the model parameters over several epochs on the target dataset.

One advantage of Mitra is its diverse set of priors, which were used during pre-training, enabling it to cover a wide range of possible data patterns. A disadvantage can be the increased computational effort, especially when fine-tuning is used.

2.4 Research question

According to current knowledge based on a scope search on pertinent literature databases, transformer models are not widely adopted for spatial location valuation and have not been directly compared with established regression and ML methods. This leads to the following research question:

- Are modern tabular transformer models suitable for predicting location values?

In a broader context, the following question also arises:

- Are tabular transformer models capable of recognising spatial structures in georeferenced datasets and generalising them for predictions?

3 Methods

3.1 Study design

These questions were answered using a use case study structured as follows (Fig. 1): First, rental price data were prepared and normalised based on apartment characteristics using hedonic price model logic (3.3). Next, multimodal accessibility to points of interest (POI) was calculated (3.4), and data on other location-describing characteristics, such as noise, population density, and impervious surface area, were prepared using Areal Weighted Interpolation (AWI) (3.5). After a Principal Component Analysis (PCA) to avoid multicollinearity between the location-describing features (3.6), traditional ML (3.7), as well as TabNet and Mitra models (3.8), were used to predict location values, and their predictive performance was compared.

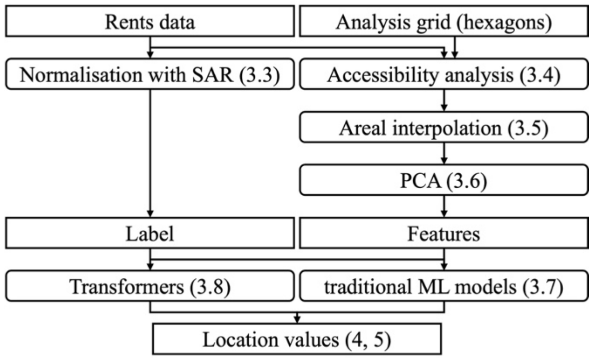

Figure 1. Methodological structure of the study.

3.2 Use case and study area

The study was conducted in Dortmund, a city in North Rhine-Westphalia (Germany) in the Ruhr area, one of Europe's largest polycentric metropolitan areas. Dortmund covers an area of approximately 280.7 km² and has a population of around 603,000 (approximately 2,150 inhabitants per km²), making it comparable to other medium-sized European cities (Destatis, 2025a). The housing market and social spaces are heterogeneous and characterised by the transformation typical of the region from coal/steel to services, logistics, and knowledge-intensive industries, comparable to the structures of other post-industrial polycentres in Europe (Hospers, 2004).

The study was conducted in close cooperation with practitioners from the city administration (including the Office for Housing, the Expert Committee for Property Values, the Surveying and Cadastral Office, and the Statistics Office) and was validated through regular feedback loops.

The data source for the location value labels was a 2022 rental survey conducted by the city administration, covering 12,400 described apartments and non-normalised rents with an arithmetic mean of ~6.62 €/m², a median of ~6.32 €/m², and a standard deviation of ~1.55 €/m². The apartment-related characteristics used for rent normalisation are summarised in Table 1; their concrete modelling and spatial regression are presented in the following Section.

Table 1. Apartment characteristics (property features) used for rent normalisation.

Category	Features
Apartment type	Basement, flat, attic, maisonette/gallery
Bathroom	No bathroom, bathroom with shower, second toilet, second bathroom with toilet
Room peculiarities	Enclosed rooms, open kitchen
Window glazing	Double-glazed windows, thermal insulation glazing, higher-quality glazing
Outdoor access	Balcony/terrace, exclusive use of garden
Floor covering	Without floor covering, laminate/carpet/simple PVC, parquet/planed floorboards, ceramic/natural stone, high-quality PVC
Heating	Tenant heating, tenant hot water, multi-apartment heating, district heating, underfloor heating
Equipment peculiarities	No intercom system, electric roller blinds
Accessibility	Barrier-free/few barriers, lift
Renovation	Renovated, bathroom renovation (2009-2014 or after 2015), flooring after 2009, insulation (1995-2014 or after 2015), electrical installation after 2009, window replacement after 2015, heating after 2009
Contract renewal	<2 years, >=2 years
Rental period in years	<1, 1-<2, 2-<3, 3-<4, 4-<6, 6-<8, 8-<10, 10-<15, 15-<20, >=20
Landlord type	Pseudonymised code (e.g., for private landlords, housing associations)
Age	Construction year
Living space	Living space in m^2

3.3 Rent normalisation

The aim of rent normalisation was to derive comparable rents from available rent data by accounting for differences in apartment characteristics. This allowed the remaining spatial differences in the normalised rents to be interpreted as location effects. This was in line with hedonic price models, in which price can be understood as a function of a set of characteristics (Rosen, 1974). Methodologically, this corresponded to the calculation of a 'constant quality' or reference/norm object: prices were converted into the features of the norm object to attribute differences in normalised rents primarily to location and environment (Sheppard, 1999; Mueller-Kett, 2025).

Prior to modelling, an automated data preparation and verification process was implemented. This included a spatial-temporal plausibility check, an error-value check, a plausibility check of the value ranges, a multicollinearity check, a test for spatial autocorrelation, and a visual plausibility check based on cartographic representations. This resulted in the removal of 18 data points.

Significant spatial autocorrelation was found in the rent data (Moran's $I \approx 0.62$, $E \approx 0$, $p \approx 0$, spatial weighting matrix based on 8 nearest neighbours). This supported the assumption that rents were spatially structured.

Consequently, spatial regression models were estimated using the property characteristics listed in Table 1 as regressors to quantify their influence. The following approaches were tested as spatial model candidates (LeSage & Pace, 2009; Pebesma & Bivand, 2023):

- Spatial Lag Model (SAR/SLM)
- Spatial Error Model (SEM)
- Spatial Durbin Model (SDM)
- Spatial Durbin Error Model (SDEM)

To reduce overfitting and test generalisability, 10-fold cross-validation was performed for each model candidate, and the Root Mean Squared Error (RMSE) was used as the target metric. Since cross-validation with a small number of folds can lead to inaccurate error estimates, a Burman correction was also applied to account for the number of folds (Burman, 1989). In addition to these metrics, the RMSE spread across the folds is shown in Table 2.

As shown in Table 2, the SDM achieved the best performance and was used for rent normalisation.

Table 2. Model performance for rent normalisation in a 10-fold cross-validation.

Modell	RMSE - uncorrected (€/m²)	RMSE – Burman-corrected (€/m²)	RMSE standard-deviation (€/m²)
SDM	**0.822**	**0.837**	**0.051**
SAR/SLM	0.881	0.897	0.062
SEM	0.891	0.912	0.062
SDEM	0.830	0.845	0.052

For normalisation, the inverse values of the relative non-spatial regression coefficients of the SDM model were used as normalisation factors and applied to the rents.

Although omitted-variable effects and stochastic noise cannot be ruled out, the result was spatially referenced, normalised rents, whose remaining differences could no longer be explained by apartment characteristics but by location differences. These normalised rents then served as the label for estimating location values in downstream ML/transformer models.

3.4 Accessibility analyses

Accessibility to Points of Interest (POI) was calculated as a key location-influencing feature. Official POI data were provided by the city administration and enriched with OpenStreetMap (OSM) data, with the official data set specifying the POI categories to be used. Since the reality of urban residents' lives is not necessarily aligned with administrative city boundaries, especially in polycentric urban regions, and to reduce edge artefacts, the study area was extended by a 1 km buffer around the city’s administrative border.

To allow for flexible adaptation and expansion of the POI inventory, it was organised so that it could be modified via a simple CSV file, an excerpt of which is shown in Table 3.

Table 3. Excerpt from the CSV file specifying the official POI categories and OSM enrichments

Category in official data set	Enrichment with OSM data
Pharmacy	amenity=pharmacy
Doctor	amenity=doctors
Long-distance network connection	highway=motorway_link
City centre	place=town
Local peculiarities and sights	-
Food	amenity=restaurant, amenity=fast_food, shop=snack
Green spaces	leisure=park, leisure=garden
Primary school	amenity=school;isced:level=1
Train main station	railway=station;name:de= Dortmund Hbf
Kindergarten	amenity=kindergarten, amenity=school;isced:level=0
Hospital	amenity=hospital
Cultural venue	leisure=music_venue, leisure=festival_grounds, amenity=exhibition_centre, amenity=cinema, amenity=stage, amenity=culture_center
Museum	tourism=museum
Groceries	shop=supermarket, shop=grocery
Local shopping area	-
Pool	amenity=swimming_pool
Playground	leisure=playground
Sports	sport=*, leisure:sports_centre
District centre	place=suburb
Secondary school	amenity=school;isced:level=2, amenity=school;isced:level=3

Four modes of transport were considered:

- walking
- cycling
- motorised private transport (MPT)
- public transport

The transport network was obtained from OSM and tailored to the extended study area (including the buffer zone) using Osmium (Topf et al., 2025) to ensure consistent routing and travel-time estimates with a manageable computational load.

For public transport accessibility, timetable data were used from a GTFS dataset (Open.NRW, n. d.). The departure time was set to 23 September 2025 at 07:30:00, with a 30-minute time window and 5 draws per minute.

To account for topographical influences (especially on walking and cycling), a digital terrain model (DTM) was loaded using the R library elevatr (Hollister & Shah, 2015). The assumed walking speed without incline was 3.6 km/h, and the cycling speed was 12 km/h.

The multimodal accessibility calculations were implemented using the R library r5r, which is based on the R5 routing kernel (Pereira et al., 2021).

For each apartment, the accessibility of all POI categories was calculated as the shortest travel times for each mode of transport. Methodologically, this corresponds to a closest facility/nearest destination logic per category (Handy & Niemeier, 1997; Geurs & van Wee, 2004).

For the subsequent area-wide inference and cartographic representation of the location values, all location-related features (including the accessibility of POIs) were calculated twice: (i) for all rental properties and (ii) for an analysis grid. The analysis grid consisted of hexagons measuring 100 m × 100 m, with accessibility calculated at each hexagon's centroid.

3.5 Areal interpolation

In addition to POI accessibility, other location-describing characteristics have been integrated:

- Population density from the 2022 nationwide census (Destatis, 2025b), spatially referenced to the INSPIRE GeoGrid of the Federal Agency for Cartography and Geodesy (BKG, 2025)
- Commercial and industrial areas (including 100-metre buffer), urban green spaces with scenic effect, and usable green spaces from land-use mapping in 2021 by the Regional Planning Association for the Ruhr Area (RVR, 2025)
- Impervious surfaces from aerial image-based derivations, generated via the ADOIS project (for more information, see Maryniak, 2025)

- Noise pollution as a 24-hour average and a night-time average, provided by the City of Dortmund (administrative data)

The selection and integration of these data sets was controlled via a CSV configuration file, analogous to the specification of the OSM data to be used. This allows city administration employees to add or update datasets without modifying the code.

These source data were available in different geometries (points, polygons, rasters) and, naturally, were not congruent with the hexagon grid, which represented the target geometry for the location values to be calculated. This meant that, in addition to purely technical issues of overlay, the Modifiable Area Unit Problem (MAUP) (Fotheringham & Wong, 1991) and a change of support (Gotway & Young, 2002) were touched upon. Although choosing a regular hexagonal grid reduced the zoning effect compared to administrative polygons, it did not eliminate these problems.

The purely technical transfer was not critical for rental properties (points), as the data could be added using spatial joins, in which the source geometries were irrelevant. For the analysis grid, however, the area data had to be transferred from incompatible zone systems to hexagons. This presented a two-zone problem in which the source and target zones only partially overlapped, so that values could not be transformed without assumptions about the spatial distribution within the source zones (Gotway & Young, 2002). Areal Weighted Interpolation (AWI) was used for harmonisation (Goodchild & Lam, 1980) using the implementation in the R library areal (Prener & Revord, 2019).

3.6 Principal component analysis

The processed location-descriptive data (POI accessibility, population density, imperviousness, etc.) served as features for subsequent model training. Since these variables often vary together conceptually and empirically (e.g., high population density and good POI accessibility), multicollinearity was to be expected, which can impair parameter estimates and interpretability, especially in regression-based models (Dormann et al., 2013). PCA was therefore used to decouple the features' structure (Jolliffe & Cadima, 2016). In addition, since the spatial coordinates were incorporated into this transformation, spatial interactions were encoded, much like positional encoding in classical transformers.

PCA is typically used in analyses for two reasons: (i) dimensionality reduction by using only the first principal components (not applied here) and (ii) transformation of the axis space in which the principal components are orthogonal to each other in pairs and thus uncorrelated, whereby the information content is minimally reduced or, when all principal components are used, remains unchanged (Jolliffe & Cadima, 2016). In this study, PCA was used as an orthogonalising transformation to avoid multicollinearity among features, and all principal components were used.

3.7 Machine learning

The h2o library for R was used to systematically train classical ML models to predict location values based on the prepared features. This library serves as a client for the H2O engine, which can train models in parallel on multi-core hardware via multi-threading and, optionally, in a cluster of multiple nodes (Fryda et al., 2024).

As part of these experiments, 100 model candidates from the following model families were trained and compared:

- Generalised Linear Model (GLM)
- Gradient Boosting Machine (GBM)
- Distributed Random Forest (DRF)/Extremely Randomised Trees (XRT)
- Deep Learning (DL): Feedforward-Networks/Multi-Layer Perceptron (MLP) for tabular data
- Stacked Ensembles: Including a meta-learner that combines several models (see below)

The H2O-supported automated grid search was used for hyperparameter tuning, which defines and tests algorithm-specific default search spaces.

In addition, model stacks were also considered. Several powerful models were trained, and their predictions were then combined by a meta-model. This method can reduce generalisation errors, as different models often exhibit distinct error profiles (Wolpert, 1992).

For each model candidate, 90% of the available data was used for training, and this subset was further split into 5 folds for cross-validation to select the model. Finally, the remaining 10% of the available data, which had been retained as test data, was used for the final evaluation.

The following metrics were calculated for model evaluation and are presented in the results section:

- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- Mean Residual Deviance (MRD)
- Root Mean Squared Error (RMSE)

3.8 Transformer models

TabNet and Mitra models were considered for estimating location values with tabular transformers. Since neither approach was consistently available for R at the time of the study with an equally stable range of functions, the modelling was implemented in Python. To support future usability in the city administration, the respective Python scripts were triggered from a uniform, central R Markdown document. To ensure comparability with classic machine learning models, identical training and test partitions (i.e., the same splits) were used.

The Python implementation in the pytorch_tabnet library was used, which provides TabNet architectures based on PyTorch (DreamQuark, 2023). Three TabNet variants with different maximum training lengths were tested: 10, 50, and 100 epochs, all trained on the same data.

Optuna was used for hyperparameter tuning, a framework for automated hyperparameter optimisation that declaratively defines search spaces and provides efficient methods, such as advanced sampling, to accelerate the search (Akiba et al., 2019). The following search spaces were considered for model training (for more information on the respective parameters, see DreamQuark, 2023):

- Dimensions per step: {16, 32, 48, 64}
- Attention layer width: {16, 32, 48, 64}
- Number of decision steps: {3, 4, 5, 6, 7, 8, 9, 10}
- Degree of feature reuse (γ): [1.0, 2.0]
- Sparsity regularisation (λ_{sparse}): [0.0001, 0.01]
- Learning rate: [0.001, 0.02]
- Weight decay: [0.000001, 0.0001]
- Mini-batch size: {512, 1024, 2048}
- Virtual batch size: {64, 128, 256}

The second transformer model tested was the Mitra model. Mitra is a tabular transformer model that follows the classic transformer pre-training/fine-tuning paradigm more closely than TabNet (Zhang et al., 2025). The Python library AutoGluon was used to model tabular data via an AutoML-oriented interface (Erickson et al., 2020). As with the other model families, the same training and test partitions were used to ensure comparability.

Two Mitra variants were considered: (i) without fine-tuning and (ii) with fine-tuning on the prepared dataset with 10 fine-tuning steps.

Analogous to the traditional machine learning models, the same error metrics were calculated and evaluated for the transformer models.

3.9 Data and Software Availability Section

The analyses were primarily implemented in R and R Markdown; POI accessibility was calculated using Java Virtual Machine (JVM)-based routines; OSM data was processed with Osmium; and tabular transformer models were used and finetuned in Python.

Since subsequent updates and further development were planned to be conducted by the city administration, data protection regulations required local data processing. Consequently, development was carried out on local hardware (MacBook Pro M4 Max with a 40-core GPU and 128 GB of RAM), with a focus on portability across machines. To this end, R environments were isolated with renv (Ushey & Wickham, 2019), and modular workflows were developed that were easily customisable and expandable via simple CSV files.

Due to privacy concerns and project-related agreements, the data cannot be published, and the developed code is not openly available. However, the developed workflow consists of open-source software, and all relevant dependencies, as well as the openly available datasets, are described in the referenced sources.

4 Results

This section presents the empirical results of the location value estimations and compares the predictive performance of the tabular transformer models with that of traditional machine learning approaches.

4.1 Model metrics

As shown in Table 4, both the Mitra models – with and without fine-tuning (FT) – and, in particular, the TabNet model with up to 100 epochs showed prediction performance comparable to that of the stacked ensemble model, consisting of traditional ML algorithms. This holds true not only for the RMSE values as the primary key metric, but also for the other calculated error measures. The traditional ML models showed the best values across all metrics, but the transformer models, especially the best TabNet model (TabNet 100), ranked closely behind. The best non-stacked, traditional ML model was a Gradient Boosting Machine (GBM) with performance metrics similar to the stacked ensemble, differing only in the fourth decimal place.

This GBM comprised a model with 41 trees and a per tree-depth of 17, ranging from 148 to 531 leaves.

Table 4. Rounded model metrics of the best-performing models.

Modell	RMSE	MSE	MAE	MRD
Stacked ensemble	**0.728**	**0.529**	**0.460**	**0.073**
Best GBM	0.727	0.529	0.462	0.073
TabNet 10	0.910	0.823	0.641	0.116
TabNet 50	0.783	0.613	0.512	0.084
TabNet 100	**0.775**	**0.601**	**0.513**	**0.083**
Mitra	0.855	0.730	0.586	0.101
Mitra FT	***0.832***	***0.693***	***0.568***	***0.096***

The best traditional ML model was a stacked ensemble comprising 16 neural networks, 11 GBMs, 1 GLM, and an additional GLM meta-learner.

The best TabNet model (TabNet 100) was trained with 99 epochs (max. 100) with 64 neurons per step, an attention layer width of 64, 7 decision steps, a feature reuse degree of $\gamma \approx 1.06$, a sparsity regulation of $\lambda_{sparse} \approx 0.0003$, a learning rate of ~0.014, a weight decay of ~2.37, a mini-batch size of 512, and a virtual batch size of 64 (see DreamQuark (2023) for more information on the respective parameters).

The pre-trained Mitra model was a 12-layer transformer with 4 attention heads, 512-dimensional embeddings, and about 72 million parameters (Zhang et al., 2025). The model was fine-tuned on the specified local hardware for about 9 days, and the resulting model used ~7.7 GB of RAM.

Figure 2 presents an additional perspective for the model comparison. The training RMSE values for the 105 models (100 H2O, 3 TabNet, 2 Mitra) are shown on the left-hand side, with only the 9 best models during training named, and the other values only indicated for clarity. On the right-hand side, the test RMSE comparisons for the best traditional ML model (stacked ensemble), TabNet, and Mitra models are shown. It demonstrates that the model stack achieved the strongest predictive performance on unseen data, while the performance of the fine-tuned Mitra and, particularly, the TabNet model with up to 100 epochs fell within a comparable range.

This figure also indicates that, as expected, the test RMSE values were higher than the training RMSE values. However, the fact that this increase cannot be described as sudden relative to the target value (mean of the normalised rents ~7.4 €/m²) suggests that overfitting cannot be ruled out but assessed as minor.

4.2 Model predictions and location values

Figure 3 shows the location values derived from the model predictions as spatial patterns for the entire study area. To this end, the predicted normalised rents were projected onto a location value scale commonly used in North Rhine-Westphalia, where the study was conducted: 1 = good to 8 = simple location. This transformation primarily served the purpose of interpretability and compatibility with established administrative and valuation practices. The resulting location values were stored alongside geometries and metadata in a GeoPackage to facilitate transfer to the practice partners' GIS workflows.

It should be noted that this figure is not intended to provide a detailed view of the calculated location values. Rather, it demonstrates that, despite some local differences, the models tested are generally capable of mapping spatial patterns consistently and reflecting this in

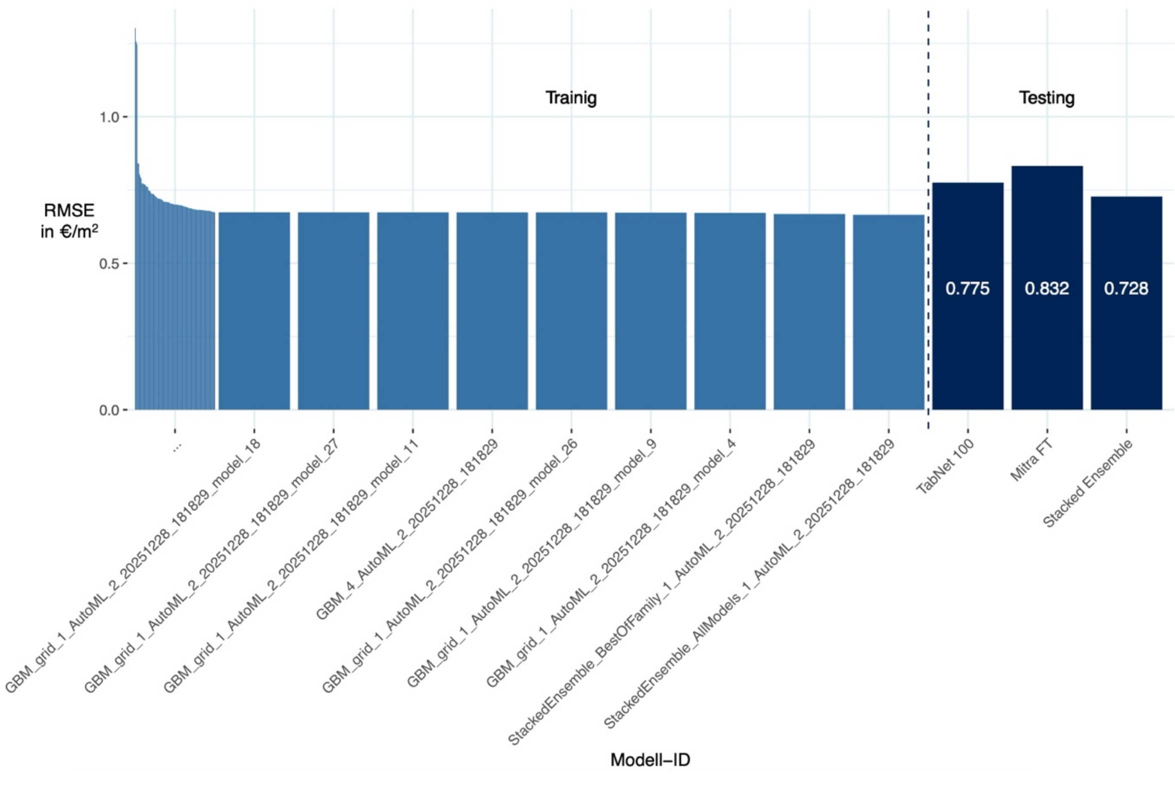

Figure 2. Comparison of the model prediction performance of the most efficient stacked ensemble, TabNet, and Mitra models.

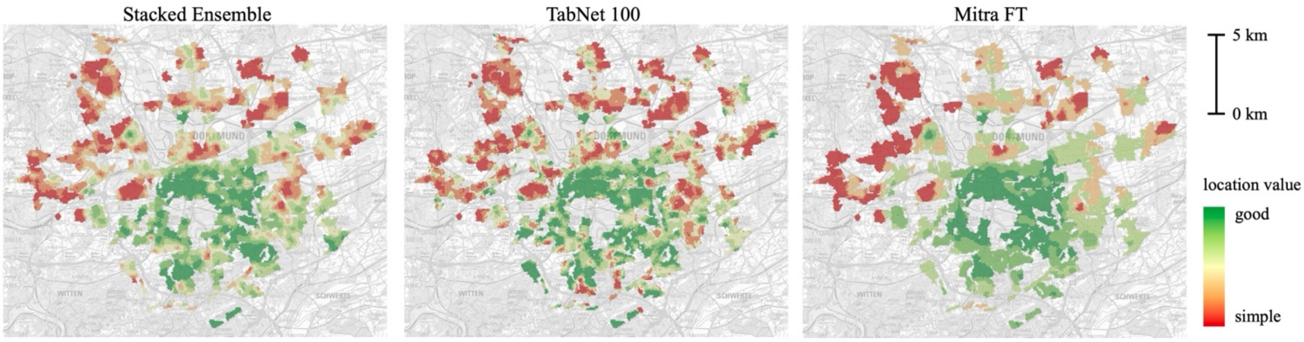

Figure 3. Schematic representation of the generated location values based on the stacked ensemble and the best performing TabNet and Mitra models (basemap: Regional government Cologne).

their predictions, with the model results showing comparable patterns. For example, all three model predictions show a clear north-south gradient and a less pronounced east-west gradient in comparable areas of the city. Similarities are also evident in the city centre and in the more detailed patterns of the outlying areas. The spatial structure in the model predictions is reflected in highly significant autocorrelation for all models (Tab. 5). Furthermore, a local cluster analysis with a BH-FDR correction (q = 0.05; Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001) revealed similar spatial patterns in the LISA clusters across all model predictions (Fig. 4). An additional comparison of the model predictions is possible by directly comparing the prediction differences (Fig. 5). These differences are generally low, particularly when comparing the stacked ensemble model and the best TabNet model.

Table 5. Rounded values of autocorrelation tests of the models' predictions.

Modell	Moran's I	Expected	z statistic	p
Stacked ensemble	0.985	0.00007	190.83	<<0.05
TabNet 100	0.962	0.00007	186.13	<<0.05
Mitra FT	0.994	0.00007	192.47	<<0.05

It should be noted that the location values presented here are not intended to be regarded as final versions for administrative practice. The values presented are aggregated by the practice partners into application-related geometries, averaging minor differences within these zones, and then approved by the relevant committees and administrative bodies. The data generated here forms the basis for this process.

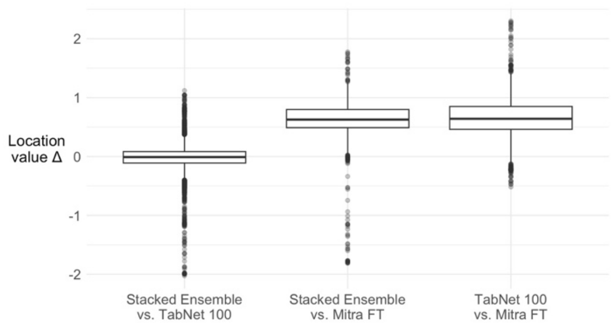

Figure 5. Differences in location values predicted by different models.

A plausibility check of the results was carried out informally through repeated interviews with seven experts from various municipal departments. The partners rated the spatial patterns as plausible overall and emphasised that they were largely consistent with their local expertise and knowledge of the area. At the same time, isolated

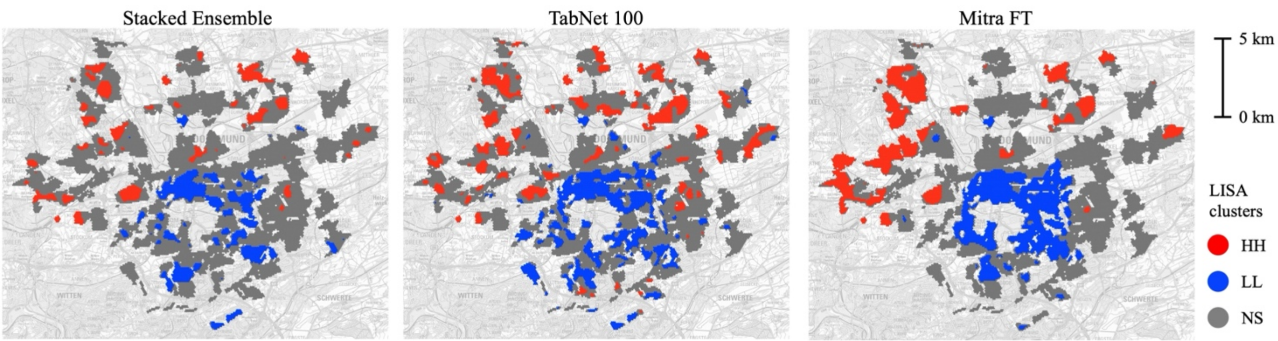

Figure 4. LISA clusters based on model predictions (basemap: Regional government Cologne).

edge effects and small-scale heterogeneities were described, requiring a model explanation (see below).

4.3 Transfer to practice

The implementation, as an R Markdown workflow, combined analysis and documentation in a consistent form, which the practice partners considered significant for applicability. In addition, a modular structure with outsourced configuration files increased maintainability and enabled even non-experts to adjust key parameters without requiring code changes. Regular exchanges with practice partners also increased the acceptance and usability of the analysis logic and results, thereby contributing significantly to the transfer of knowledge into practice.

Since the models used are often black boxes, model explainability was necessary for the acceptance and use of the results. To this end, SHAP was used as a model-agnostic approach, but this required substantial computational effort (Lundberg & Lee, 2017). In addition, for the fine-tuned Mitra and the TabNet models, feature importance was used for explainability, as provided by the AutoGluon and pytorch_tabnet implementations. This increased interpretability, but the explained features were the principal components, and the content interpretation then had to be traced back to the original features via the PC loadings (Jolliffe & Cadima, 2016), stating a trade-off between model interpretability and predictive performance. Nevertheless, practitioners considered this step towards model explainability to be helpful and important for transfer to administrative practice.

5 Discussion

Although the tabular transformer models were still outperformed by (stacked) traditional ML models in this case study, they demonstrated comparable performance and were well-suited to estimating plausible location values. At the same time, the methodology presented has several limitations, which will be discussed in this section.

First, when using area-based data, the MAUP problem, change-of-support, and the two-zone problem were encountered. In this study, these are recognised as methodological restrictions, but are considered acceptable due to (i) relatively small source zones, (ii) the semantically appropriate and transparent AWI transfer, and (iii) the consistent transfer to a uniform analysis grid that was used across all models. The last point, in particular, is considered crucial, as the analytical focus of this study is on model comparison, and all models considered were affected in the same way.

One practical limitation concerns the resources required for training: tabular transformers are more computationally intensive than many classic ML models. Training the fine-tuned Mitra model took ~9 days in this study, compared to ~190 minutes for training all traditional ML models. The training performance of the transformer models could likely be improved with more powerful hardware or cloud-based clusters, but this would conflict with their applicability in municipal practice and the associated regulations. This would also allow for expansion of the hyperparameter search space. For example, the parameters found for the TabNet model were mostly at the edge of the search space, suggesting that an expansion could improve the model's predictive performance.

Another limitation is treating location as a purely monetary factor. This perspective is plausible in many applications, especially in value and risk assessment, because the real estate market is highly relevant to macroeconomics and financial stability, and prices therefore directly guide action in these fields (Jordà et al., 2016). At the same time, the calculated location values are intended for use beyond these fields of application, in which non-monetary dimensions, such as ecological and cultural values, are ignored as model target variables. The practicality of this approach is supported by the fact that willingness to pay partially capitalises on such non-material qualities: When people are willing to pay more for 'soft' location qualities such as tranquillity, green spaces, or cultural importance, this is often reflected, at least in part, in higher rents. Market prices can therefore serve as a rough, albeit imperfect, aggregate indicator of various value dimensions (Rosen, 1974).

Furthermore, there is a market assumption that the observed rents reflect the demand side's preferences under sufficiently competitive conditions. However, this assumption can be questioned, in part, e.g., by regulations and allocation mechanisms, such as social housing. Nevertheless, according to the practice partners, the dataset used was sufficiently adjusted for this factor, e.g., by excluding publicly subsidised properties, including those subject to rent controls.

Finally, rapid methodological developments in machine learning, and transformer models in particular, are a limitation in their own right. New architectures and training recipes emerge quickly, while practical transfer is often delayed or fails due to 'technical debt' (Sculley et al., 2015). Nevertheless, this study remains practical if the contribution is not understood as the 'final word' on the best architecture, but rather as robust evidence that tabular transformer-based models (i) work in the given urban analytical setup, (ii) can satisfactorily map patterns in spatial data sets, and (iii) can be embedded in modular pipelines in such a way that updates are possible with reasonable effort in local city administrations.

6 Conclusion

In this study, modern tabular transformers were used to derive and spatially map location values from normalised rents and location-descriptive features across the study area. The key finding is that these models not only achieve predictive performance comparable to established machine learning methods but also reproduce spatial patterns that are plausible relative to traditional ML models.

As measured by RMSE, a model stack combining several classical ML models demonstrated the strongest overall performance, alongside a GBM that performed almost as well. Nevertheless, it should be emphasised that the transformers, particularly TabNet with 100 epochs, achieved performance comparable to that of the stacked ensemble and the GBM.

Overall, the study demonstrates that transformer models expand the methodological toolkit for spatial analysis. Given the limitations identified in the discussion section, these can be understood as concrete starting points for further research to advance the use of modern transformer models in spatial data analysis.

Declaration of Generative AI in writing

The author declares that he has used Generative AI tools in the preparation of this manuscript. Specifically, the AI tools were only utilised for language editing, improving grammar, and sentence structure, but not for generating scientific content, research data, or substantive conclusions. All intellectual and creative work, including the analysis and interpretation of data, is original and has been conducted by the author without AI assistance.

References

Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., and Oluwadara, G.: House price prediction using random forest machine learning technique, Procedia Computer Science, 199, 806–813, https://doi.org/10.1016/j.procs.2022.01.100, 2022.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A next-generation hyperparameter optimization framework, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '19), https://doi.org/10.1145/3292500.3330701, 2623–2631, 2019.

Anselin, L.: Spatial econometrics: methods and models, Kluwer Academic Publishers, Dordrecht, Netherlands, 1988.

Arik, S. Ö. and Pfister, T.: TabNet: Attentive Interpretable Tabular Learning, Proceedings of the AAAI Conference on Artificial Intelligence, 35, 6679–6687, https://doi.org/10.1609/aaai.v35i8.16826, 2021.

Basu, S. and Thibodeau, T. G.: Analysis of spatial autocorrelation in house prices, The Journal of Real Estate Finance and Economics, 17, 61–85, https://doi.org/10.1023/A:1007703229507, 1998.

Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing, Journal of the Royal Statistical Society, 57, 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x, 1995.

Benjamini, Y. and Yekutieli, D.: The control of the false discovery rate in multiple testing under dependency, The Annals of Statistics, 29, 1165–1188, https://doi.org/10.1214/aos/1013699998, 2001.

Breiman, L.: Statistical modeling: The two cultures (with comments and a rejoinder by the author), Statistical Science, 16, 199–231, https://doi.org/10.1214/ss/1009213726, 2001a.

Breiman, L.: Random forests. Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001b.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, … Amodei, D.: Language Models are Few-Shot Learners, Advances in Neural Information Processing Systems, 33, 1877–1901, 2020.

Bürgerliches Gesetzbuch (BGB): § 558d Qualifizierter Mietspiegel [Qualified rent index], available at: https://www.gesetze-im-internet.de/bgb/__558d.html, last access: 9 January 2026.

Burman, P.: A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods, Biometrika, 76, 503–514, https://doi.org/10.1093/biomet/76.3.503, 1989.

Can, A.: Specification and estimation of hedonic housing price models, Regional Science and Urban Economics, 22, 453–474, https://doi.org/10.1016/0166-0462(92)90039-4, 1992.

Cortes, C. and Vapnik, V.: Support-vector networks, Machine Learning, 20, 273–297, https://doi.org/10.1007/BF00994018, 1995.

Day, B., Bateman, I. J., Lake, I., and Lovett, A. A.: Beyond implicit prices: Recovering theoretically consistent and transferable values for noise avoidance from a hedonic property price model, Environmental and Resource Economics, 37, 211–232, https://doi.org/10.1007/s10640-007-9121-8, 2007.

Debrezion, G., Pels, E., and Rietveld, P.: The impact of railway stations on residential and commercial property value: A meta-analysis, The Journal of Real Estate Finance and Economics, 35, 161–180, https://doi.org/10.1007/s11146-007-9032-z, 2007.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 4171–4186, https://doi.org/10.18653/v1/N19-1423, 2019.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., … , and Lautenbach, S.: Collinearity: A review of methods to deal with it and a simulation study evaluating their performance, Ecography, 36, 27–46, https://doi.org/10.1111/j.1600-0587.2012.07348.x, 2013.

DreamQuark: TabNet: Attentive interpretable tabular learning (Version 4.1.0) [Software], available at: https://github.com/dreamquark-ai/tabnet, last access: 9 January 2026, 2023.

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., and Smola, A.: AutoGluon-Tabular: Robust and accurate AutoML for structured data, arXiv, https://doi.org/10.48550/arXiv.2003.06505, 2020.

Federal Agency for Cartography and Geodesy (BKG): Geographische Gitter für Deutschland in Lambert-Projektion [Geographical grid for Germany in Lambert-Projection] (GeoGrid Inspire) [Dataset], available at: https://gdz.bkg.bund.de/index.php/default/inspire/sonstige-inspire-themen/geographische-gitter-fur-deutschland-in-lambert-projektion-geogitter-inspire.html, last access: 9 January 2026, 2025.

Federal Statistical Office Germany (Destatis): Städte nach Fläche, Bevölkerung und Bevölkerungsdichte am 31.12.2024 [Cities by area, population and population density on 31.12.2024] [Dataset], available at: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/05-staedte.html?templateQueryString=Bevölkerung+Dortmund, last access: 9 January 2026, 2025a.

Federal Statistical Office Germany (Destatis): Ergebnisse des Zensus 2022 [Results of the 2022 Census] [Dataset], available at: https://www.destatis.de/EN/Themes/Society-Environment/Population/Census2022/_node.html, last access: 9 January 2026, 2025b.

Friedman, J. H.: Greedy function approximation: A gradient boosting machine, The Annals of Statistics, 29, 1189–1232, https://doi.org/10.1214/aos/1013203451, 2001.

Fotheringham, A. S. and Wong, D. W. S.: The modifiable areal unit problem in multivariate statistical analysis, Environment and Planning A, 23, 1025–1044, https://doi.org/10.1068/a231025, 1991.

Fryda, T., LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., …, and Wong, W.: h2o: R Interface for the 'H2O' Scalable Machine Learning Platform [R package], available at: https://doi.org/10.32614/CRAN.package.h2o, last access: 9 January 2026, 2024.

Geurs, K. T. and van Wee, B.: Accessibility evaluation of land-use and transport strategies: Review and research directions, Journal of Transport Geography, 12, 127–140, https://doi.org/10.1016/j.jtrangeo.2003.10.005, 2004.

Gibbons, S. and Machin, S.: Valuing rail access using transport innovations, Journal of Urban Economics, 57, 148–169, https://doi.org/10.1016/j.jue.2004.10.002, 2005.

Goodchild, M. F. and Lam, N. S. N.: Areal interpolation: A variant of the traditional spatial problem, Geo-processing, 1, 297–312, 1980.

Gorishniy, Y., Rubachev, I., Khrulkov, V. and Babenko, A.: Revisiting deep learning models for tabular data, Advances in Neural Information Processing Systems, 34, 18932–18943, 2021.

Gotway, C. A. and Young, L. J.: Combining incompatible spatial data, Journal of the American Statistical Association, 97, 632–648, https://doi.org/10.1198/016214502760047140, 2002.

Handy, S. L. and Niemeier, D. A.: Measuring accessibility: an exploration of issues and alternatives, Environment and Planning A: Economy and Space, 29, 1175–1194, 1997.

Hollister, J., Shah, T., Nowosad, J., Robitaille, A., Beck, M., and Johnson, M.: elevatr: Access Elevation Data from Various APIs [R package], available at: https://doi.org/10.5281/zenodo.8335450, last access 9 January 2026, 2025.

Hospers, G. J.: Restructuring Europe's rustbelt: The case of the German Ruhrgebiet, Intereconomics, 39, 147–156, 2004.

Huang, X., Khetan, A., Cvitkovic, M. and Karnin, Z.: TabTransformer: Tabular Data Modeling Using Contextual Embeddings, arXiv, https://doi.org/10.48550/arxiv.2012.06678, 2020.

Immobilienwertermittlungsverordnung [ImmoWertV]: Verordnung über die Grundsätze für die Ermittlung der Verkehrswerte von Grundstücken [Ordinance on the principles for determining the market values of real estate] (BGBl. I, p. 2805), available at: https://www.gesetze-im-internet.de/immowertv_2022/BJNR280500021.html, last access: 9 January 2026, 2021.

Jefatura del Estado: Ley 12/2023, de 24 de mayo, por el derecho a la vivienda [Law 12/2023 of 24 May on the right to housing] (BOE-A-2023-12203), Boletín Oficial del Estado, available at: https://www.boe.es/buscar/act.php?id=BOE-A-2023-12203, last access: 9 January 2026, 2023.

Jolliffe, I. T. and Cadima, J.: Principal component analysis: A review and recent developments, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374, 20150202, https://doi.org/10.1098/rsta.2015.0202, 2016.

Jordà, Ò., Schularick, M. and Taylor, A. M.: The great mortgaging: Housing finance, crises, and business cycles, Economic Policy, 31, 107–152, https://doi.org/10.1093/epolic/eiv017, 2016.

Lundberg, S. M. and Lee, S. I.: A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems, 30, 2017.

Maryniak, M.: Automatic detection of impervious surfaces (ADOIS) [Software], available at: https://github.com/geospaitial-lab/adois/tree/main, last access: 9 January 2026, 2025.

Mietspiegelverordnung (MsV): § 19 Wohnlagen [Location values], available at: https://www.gesetze-im-internet.de/msv/__19.html, last access: 27 March 2026.

Moghimi, F., Johnson, R. A., and Krause, A.: Rethinking Real Estate Pricing with Transformer Graph Neural Networks (T-GNN), 2023 International Conference on Machine Learning and Applications, 1405–1411, https://doi.org/10.1109/ICMLA58977.2023.00212, 2023.

Moreno, C., Allam, Z., Chabaud, D., Gall, C., and Pratlong, F.: Introducing the “15-minute city”: Sustainability, resilience and place identity in future post-pandemic cities, Smart Cities, 4, 93–111, https://doi.org/10.3390/smartcities4010006, 2021.

Mueller-Kett, C.: Housing location values and beyond: Towards data-driven distillation of space using artificial intelligence, 19th International Conference on Computational Urban Planning and Urban Management (CUPUM 2025), London, UK, https://doi.org/10.17605/OSF.IO/ABYQH, 2025.

Open.NRW: Soll-Fahrplandaten NRW (ÖPNV) [target public transport schedule data] [Dataset], available at: https://open.nrw/dataset/soll-fahrplandaten-nrw-oepnv, last access: 9 January 2026.

Panduro, T. E. and Veie, K. L.: Classification and valuation of urban green spaces—A hedonic house price valuation, Landscape and Urban Planning, 120, 119–128, https://doi.org/10.1016/j.landurbplan.2013.08.009, 2013.

Park, B. and Bae, J. K.: Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data, Expert Systems with Applications, 42, 2928–2934, https://doi.org/10.1016/j.eswa.2014.11.040, 2015.

Pebesma, E. and Bivand, R.: Spatial Data Science: With Applications in R, Chapman and Hall/CRC, Boca Raton, https://doi.org/10.1201/9780429459016, 2023.

Pereira, R. H. M., Schwanen, T. and Banister, D.: Distributive justice and equity in transportation, Transport Reviews, 37, 170–191, https://doi.org/10.1080/01441647.2016.1257660, 2017.

Pereira, R. H. M., Saraiva, M., Herszenhut, D., Braga, C. K. V., and Conway, M. W.: r5r: Rapid Realistic Routing on Multimodal Transport Networks with R5, R Findings, https://doi.org/10.32866/001c.21262, 2021.

Prener, C. G. and Revord, C. K.: areal: An R package for areal weighted interpolation, Journal of Open Source Software, 4, https://doi.org/10.21105/joss.01221, 2019.

Regional government Cologne: Digitale Topographische Karte 1:100 000 [WMS basemap], available at: https://www.wms.nrw.de/geobasis/wms_nw_dtk100, last access: 9 January 2026.

Regionalverband Ruhr (RVR): Flächennutzungskartierung [Land use mapping] [Dataset], available at: https://www.rvr.ruhr/daten-digitales/geodaten/flaechennutzungskartierung/, last access: 9 January 2026, 2025.

Republic of Austria: Bundesgesetz vom 12. November 1981 über das Mietrecht (Mietrechtsgesetz – MRG) [Federal Act of 12 November 1981 on tenancy law (Tenancy Act—MRG)], available at:

https://www.ris.bka.gv.at/GeltendeFassung.wxe?pageId=7643&Abfrage=Bundesnormen&Gesetzesnummer=10002531, last access: 9 January 2026, 1981.

Ribeiro, M. T., Singh, S. and Guestrin, C.: "Why should I trust you?" Explaining the predictions of any classifier, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144, https://doi.org/10.1145/2939672.2939778, 2016.

Rosen, S.: Hedonic prices and implicit markets: Product differentiation in pure competition, Journal of Political Economy, 82, 34–55, https://doi.org/10.1086/260169, 1974.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., …, and Dennison, D.: Hidden technical debt in machine learning systems, Advances in Neural Information Processing Systems, 28, 2015.

Sheppard, S.: Hedonic analysis of housing markets, in: Handbook of Regional and Urban Economics, 3, 1595–1635, Elsevier, Amsterdam, Netherlands, 1999.

Selim, H.: Determinants of house prices in Turkey: Hedonic regression versus artificial neural network, Expert Systems with Applications, 36, 2843–2852, https://doi.org/10.1016/j.eswa.2008.01.044, 2009.

Somepalli, G., Goldblum, M., Schwarzschild, A., Bruss, C. B., and Goldstein, T.: SAINT: Improved Neural Networks for Tabular Data via Row Attention and Contrastive Pre-Training, arxiv, https://doi.org/10.48550/arxiv.2106.01342, 2021.

Topf, J., Nakaner, Skowron, T., Berg, W., Couwenberg, W., …, and Hughes, T.: osmium-tool [Software], available at: https://github.com/osmcode/osmium-tool, last access: 9 January 2026, 2025.

Ushey, K. and Wickham, H.: renv: Project Environments [R package], available at: https://doi.org/10.32614/CRAN.package.renv, last access: 9 January 2026, 2019.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.: Attention is All You Need, Advances in Neural Information Processing Systems, 30, 2017.

Wolpert, D. H.: Stacked generalization, Neural Networks, 5, 241–259, https://doi.org/10.1016/S0893-6080(05)80023-1, 1992.

Zhang, X., Maddix, D. C., Yin, J., Erickson, N., Ansari, A. F., …, and Wang, B.: Mitra: Mixed Synthetic Priors for Enhancing Tabular Foundation Models, arxiv, https://doi.org/10.48550/arxiv.2510.21204, 2025.

Zhao, Y., Chetty, G., and Tran, D.: Real Eastate Price Prediction on GenerativeLanguage Models, in: 2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, https://doi.org/10.1109/CSDE59766.2023.10487658, 2023.