Band 107

Moritz Stang

Real Estate Valuation in the Age of Artificial Intelligence – Modern Machine Learning Algorithms and their Application in Property Appraisal

International Real Estate Business School
Universität Regensburg

Schriften zu Immobilienökonomie und Immobilienrecht

Herausgeber:
IREIBS International Real Estate Business School

Prof. Dr. Sven Bienert
Prof. Dr. Stephan Bone-Winkel
Prof. Dr. Kristof Dascher
Prof. Dr. Dr. Herbert Grziwotz
Prof. Dr. Tobias Just
Prof. Gabriel Lee, Ph. D.
Prof. Dr. Kurt Klein
Prof. Dr. Jürgen Kühling, LL.M.
Prof. Dr. Gerit Mannsen
Prof. Dr. Dr. h.c. Joachim Möller
Prof. Dr. Karl-Werner Schulte HonRICS
Prof. Dr. Wolfgang Schäfers
Prof. Dr. Steffen Sebastian
Prof. Dr. Wolfgang Servatius
Prof. Dr. Frank Stellmann
Prof. Dr. Martin Wentz

Moritz Stang

Real Estate Valuation in the Age of Artificial Intelligence – Modern Machine Learning Algorithms and their Application in Property Appraisal

Die Deutsche Bibliothek – CIP Einheitsaufnahme
Moritz Stang
Real Estate Valuation in the Age of Artificial Intelligence – Modern Machine Learning Algorithms and their Application in Property Appraisal
Regensburg: Universitätsbibliothek Regensburg 2023
(Schriften zu Immobilienökonomie und Immobilienrecht; Bd. 107)
Zugl.: Regensburg, Univ. Regensburg, Diss., 2023
978-3-88246-489-4

978-3-88246-489-4
© IRE|BS International Real Estate Business School, Universität Regensburg
Verlag: Universitätsbibliothek Regensburg, Regensburg 2023
Zugleich: Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften, eingereicht an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg
Tag der mündlichen Prüfung: 20. Juli 2023
Berichterstatter: Prof. Dr. Wolfgang Schäfers
Prof. Dr. Tobias Just

Acknowledgments

Without the support of various people, this dissertation in its current form would not have been possible. Therefore, I would like to take the opportunity to express my gratitude to these individuals.

First and foremost, I would like to express my appreciation to my supervisor, Prof. Dr. Wolfgang Schäfers. I thank him for his unwavering support, the inspiring discussions and the ongoing opportunities for growth and development during the last years. The collaboration with him has been immensely beneficial both academically and personally for me. Additionally, I would like to thank Prof. Dr. Tobias Just for his support, valuable comments, and taking over as secondary supervisor.

Next, I would like to thank my co-authors Bastian Krämer, Dr. Cathrine Nagl, Prof. Dr. Wolfgang Schäfers, and PD Dr. Marcelo Cajias. Without them, completing and publishing the individual papers would not have been possible. In particular, I would like to express my sincere gratitude to Bastian Krämer for his valuable contribution to the completion of this dissertation. His scientific and personal intellect has greatly assisted me throughout this journey. I will always cherish the hours-long working sessions and extensive discussions we had together.

A special thanks also goes to my colleagues at the IRE|BS Institute, especially Bastian Krämer, Daniel Promper, Benedict von Ahlefeldt-Dehn, Nino Paulus, Jürgen Deppner, and Dr. Cathrine Nagl. Each and every one of you has shown me over the past three years what it means to work in an ambitious team. Not only professionally, but also personally, you have helped me greatly, for which I am very grateful.

From the bottom of my heart, I also want to thank my family and friends for their support. Specifically, I would like to thank my father Dr. Joachim Stang, my mother Birgit Schabesberger-Stang, and my sister Marlene Stang. Not only during the period of my dissertation, but also since my birth, I can always count on your support. I am grateful for your relentless guidance, which has played an integral role in shaping me into the person I am today. Thank you for always being there for me.

Last but not least, I would like to express my heartfelt gratitude to the partner by my side, Verena Lehner. Dear Verena, words cannot express how grateful I am for all your support and love. Without you, the past years would have been only half as fulfilling and worthwhile. I look forward to a joint future with you! The world belongs to both of us and the rest of our family!

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivation and Background

The real estate valuation sector represents one of the most important domains of the real estate industry. Valuations are provided for various reasons and are often even required by regulation. Whether purchasing, mortgaging, or insuring a property, a valuation is always an elementary part of the procedure. From an economic point of view, a correct valuation is also an important aspect that should not be neglected. In particular, the Global Financial Crisis (GFC) of 2008 showed that a combination of overheated residential real estate markets, rising interest rates on predominantly floating-rate real estate loans and exuberant real estate valuations can cause the entire global financial and economic system to collapse. It is not without reason that regulatory requirements for the valuation of real estate were tightened worldwide in the aftermath of the crisis (Mishkin, 2011). Beginning after the GFC, the appraiser industry has undergone a process of continuous change that persists to this day. This transformation has been driven by evolving regulatory requirements, growing industry-specific challenges, and the general impact of digitalization. These drivers have prompted significant shifts in the way appraisers operate and are expected to continue shaping the industry in the future. In response to these factors, the global focus has shifted towards cost-cutting and efficiency-enhancing measures. This has created a need for the adoption of innovative technologies and methodologies to optimize the valuation process and ensure accurate and reliable results.

As the Royal Institution of Chartered Surveyors (2017) shows in their report “The Future of Valuations”, the valuation process is changing significantly and essentially becomes more automated through the use of new technologies, such as Artificial Intelligence (AI) or the Blockchain technology. What is often sold as a bogeyman and with the loss of jobs, however, is necessary for the survival of the appraiser industry. The number of appraisers has declined sharply over the past few years and there are hardly any new entrants to the profession. In the USA, for example, the number of accredited appraisers declined by about 20% between 2007 and 2015 (Coyle, 2015). This trend is likely to continue in the coming years, triggering a shortage of supply of valuation service providers.

In order to overcome this lack of supply in the long term, the use of AI solutions in particular appears to be a promising problem solver. Basically, AI is a branch of computer science that aims to transform computers into intelligent systems. Through computational training and advanced algorithms, AI and its subcategory machine learning can be used to build systems that are able to perform human tasks in an automated way. Machine learning algorithms are computational processes that use data to achieve a given task without being programmed to produce a particular outcome. They are able to adapt their architecture through repetition so that they become better and better over time (El Naqa & Murphy, 2015).

Machine learning algorithms have been applied successfully in various areas ranging from spacecraft engineering to finance. Also, the application within the real estate industry and especially within the profession of real estate appraisers is discussed more and more (see, e.g., Baldominos et al., 2018). However, upon closer examination beyond academia, it becomes evident that the use of modern machine learning algorithms in the appraiser sector is not yet widespread. When automated property valuations are carried out, the underlying Automated Valuation Models (AVMs) typically rely on the logic of traditional appraiser methods or simpler parametric or semi-parametric econometric models. While initial studies already show that modern machine learning methods are capable of achieving more accurate valuation results due to their ability to capture non-linearities and joint effects (see, e.g., Kok et al., 2017; Sangani et al., 2017; Singh et al., 2020; Pace & Hayunga, 2020 and Tchuente & Nyawa, 2021), they are generally not included in the circle of applicable models for regulatory purposes. The main reason for this is the so-called "black box" image of modern machine learning algorithms (see, e.g., Adadi & Berrada, 2018). Due to the way they work, their results are not intrinsically interpretable, unlike, for example, the results of parametric or semi-parametric econometric models. This is seen as a major disadvantage and has prevented their regulatory acceptance and therefore widespread use to date. What is ignored here, however, is the fact that there are now methods that can be subsumed under the term eXplainable AI (XAI), which allow us to open the "black-box" and provide the necessary explainability (Molnar, 2020). In relation to the real estate industry and real estate research, however, such approaches have hardly been applied to date.

The fact that these methodological innovations have not yet been applied to real estate-related topics, and furthermore, that the use of machine learning methodologies in the appraisal practice is not yet widespread and often prohibited by regulations, shows, that there is still a need for theoretical research on the optimal use of modern machine learning algorithms for appraiser purposes. It appears that there is still a lack of understanding regarding the functionality of these algorithms and the interpretation of their outputs.

To improve the overall comprehension of modern machine learning algorithms and to overcome current obstacles hindering their application within the appraiser industry, this dissertation aims to explore unanswered questions that need to be addressed to fully utilize their potential in the upcoming years. Specifically, the research of this dissertation focuses on the optimal application of machine learning algorithms in the field of residential real estate valuation, their explainability, and how they can be used to determine the quality of real estate locations in an automated way. Through three distinct papers, this research will provide valuable insights for real estate researchers and practitioners, enabling them to effectively utilize modern machine learning algorithms, make more in-depth decisions, and ultimately enhance their efficiency.

The purpose of the first paper is to deliver further insights into the optimal use of modern machine learning algorithms for AVMs from both theoretical and practical perspectives. It compares these algorithms with traditional econometric and non-econometric models and performs a nationwide comparison of different AVM algorithms for the first time. Within the paper, a unique dataset with 1,212,546 observations, distributed throughout Germany, is used. The paper aims to answer critical questions related to the ongoing debate surrounding the approval of machine learning based AVMs and promote their adoption in real estate valuation practices. The results show that the applied modern machine learning approach is able to achieve the highest overall accuracy in the valuation of standard residential properties in Germany. But the results also show that for designing an AVM, there is no “one size fits all”. Although the modern machine learning approach is the best performer across the country, there are also parts in Germany where the other more traditional models are best suited for estimating market values. In this context, it is particularly evident that the respective data availability seems to play a role. The paper contains important theoretical and practical implications that will help to optimize the use of AVMs and increase their acceptance.

The second paper deals with the topic of the explainability of modern machine learning models. A non-parametric machine learning eXtreme Gradient Boosting (XGBoost) approach in combination with two XAI techniques called Permutation Feature Importance (PFI) and Accumulated Local Effects Plots (ALE) is used to show how the results of a modern machine learning model can be explained. In a real estate context, so far, XAI approaches have been explored only to a limited extent, therefore the objective of this paper is to showcase how XAI methods can be used to make the deep hidden patterns of real estate markets interpretable for human beings. The analysis covers the residential real estate markets of the seven largest cities of Germany and is carried out based on a dataset of 81,166 observations. To detect differences between different subtypes of residential properties, the dataset is further split into two groups. The first consists of around 61,763 condominiums and the second of 19,403 single-family homes. The results of the paper reveal that the same value-determining features play an important role for both condominiums and single-family homes. However, there are fundamental differences within the two property types regarding the individual influence of the respective features on the market value of a property. Especially, non-linear relationships are identified for the majority of features. In summary, the results show how important it is for both real estate research and practice to conduct data-driven analyses with the help of modern machine learning and XAI approaches, in order to gain important market insights and, if necessary, to update long-established assumptions regarding the determinants of real estate market values.

In the third and final paper, a new data-driven and automated approach to evaluate real estate locations, called "SHAP Location Score" (SHAP-LS), is introduced. This approach is based on a state-of-the-art machine learning model and a post-hoc model agnostic explainable artificial intelligence (XAI) approach. The SHAP-LS combines hedonic pricing theory and modern machine learning algorithms to provide a non-biased method for assessing the quality of a property location. The approach is characterized by its high degree of flexibility and can be implemented in a model-agnostic manner for any machine learning algorithm and for any feature set. To demonstrate the efficacy of the SHAP-LS, the approach is applied to a dataset of 26,860 residential rental properties in the city of London (UK). The results show that the approach is able to extract the value - and thus also location quality - determining factors of real estate prices on a granular level and thus to show where within the city the quality of a location is particularly good and where it is not. The findings also highlight that there are significant differences with regard to the individual willingness to pay for different levels of location quality and that this willingness seems to be highest in the central locations of the city.

1.2 Research Questions

The following section presents the research questions formulated in the context of the three individual papers. All questions deal with the topic of real estate valuation and specifically with the use of modern machine learning algorithms in this field.

Paper 1: From Human Business to Machine Learning – Methods for Automating Real Estate Appraisals and their Practical Implications

- Do machine learning methods outperform well-established AVM methods like the OLS, the GAM and the EXF, and should they therefore also be considered within the regulatory discussion of AVMs?
- Should AVMs rely on the use of one single approach, or should multiple models be integrated for different spatial areas?
- Does the performance of the methods depend on data availability and structure?

Paper 2: Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values

- How can the "black box" label of modern machine learning techniques be overcome to ensure traceability, auditability, robustness, and resilience of inputs and outputs for regulatory purposes?
- Which features are important for the market values of residential properties?
- To what extent are these features characterized by either linearity or non-linearity? Are there differences depending on different cities?
- Can fundamental differences between condominiums and single-family homes be observed?

Paper 3: Changing the Location Game – Improving Location Analytics with the Help of Explainable AI

- What kind of theoretical foundation is required to design a machine learning-based approach for a data-driven real estate location valuation?
- Can modern machine learning algorithms in combination with XAI methods be used to provide a purely data-driven method to evaluate the quality of real estate locations?
- To what extent can modern machine learning algorithms capture the complex interactions and multilayered non-linear relationships that characterize the quality of real estate locations?

1.3 Co-Authors, Submissions and Conference Presentations

The following overview provides information about co-authors, journal submissions, publication status and conferences presentations.

Paper 1: From Human Business to Machine Learning – Methods for Automating Real Estate Appraisals and their Practical Implications

Authors:

Moritz Stang, Bastian Krämer, Dr. Cathrine Nagl, Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: German Journal of Real Estate Research
Current Status: accepted (04/07/2022) and pre-published online (19/07/2022)

Conference Presentations:
This paper was presented at:

- the 38th Annual Conference of the American Real Estate Society (ARES) in Bonita Springs, USA (2022)
- the 28th Annual Conference of the European Real Estate Society (ERES) in Milan, Italy (2022)
- the 4th Workshop "Artificial Intelligence and Finance" of the Center of Finance of the University of Regensburg in Regensburg, Germany (2022)

Paper 2: Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values

Authors:

Bastian Krämer, Dr. Cathrine Nagl, Moritz Stang, Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: Journal of Housing Research
Current Status: accepted (11/01/2023) and pre-published online (14/02/2023)

Conference Presentations:
This paper was presented at:

- the 38th Annual Conference of the American Real Estate Society (ARES) in Bonita Springs, USA (2022)
- the 28th Annual Conference of the European Real Estate Society (ERES) in Milan, Italy (2022)
- the Doctoral Seminar of the Center of Finance of the University of Regensburg in Regensburg, Germany (2022)

Awards:
This paper was awarded the "Best PhD Paper Award 2022" at the 28th Annual Conference of the European Real Estate Society (ERES) in Milan, Italy (2022).

Paper 3:	Changing the Location Game – Improving Location Analytics with the Help of Explainable AI

Authors:

Moritz Stang, Bastian Krämer, PD Dr. Marcelo Cajias, Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: Journal of Real Estate Research
Current Status: Under review (17/01/2023)

Conference Presentations:
This paper was presented at:

- the 39th Annual Conference of the American Real Estate Society (ARES) in San Antonio, USA (2023)
- the Doctoral Seminar of the Center of Finance of the University of Regensburg in Regensburg, Germany (2023)

This paper will be presented at:

- the 29th Annual Conference of the European Real Estate Society (ERES) in London, UK (2023)

1.4 References

Adadi, A. and Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160.

Baldominos, A., Blanco, I., Moreno, A., Iturrarte, R., Bernárdez, Ó. & Afonso, C. (2018). Identifying Real Estate Opportunities Using Machine Learning. *Applied Sciences*, 8(11), 2321.

Coyle, M.L. (2015). *Future of Valuation*. Working RE magazine, Editor's note, San Diego.

El Naqa, I. & Murphy, M.J. (2015). *What Is Machine Learning?*. In Machine Learning in Radiation Oncology, Cham: Springer, 3–11.

Kok, N., Koponen, E.-L. & Martínez-Barbosa, C.A. (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal of Portfolio Management*, 43(6), 202–211.

Mishkin, F.S. (2011). Over the Cliff: From the Subprime to the Global Financial Crisis. *Journal of Economic Perspectives*, 25(1), 49–70.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

Pace, R.K. & Hayunga, D. (2020). Examining the Information Content of Residuals from Hedonic and Spatial Models Using Trees and Forests. *The Journal of Real Estate Finance and Economics*, 60(1-2), 170–180.

Royal Institution of Chartered Surveyors (2017). *The Future of Valuations: The relevance of real estate valuations for institutional investors and banks – views from a European expert group*. RICS, London.

Sangani, D., Erickson, K. & Hasan, M.A. (2017). *Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting*. In IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Orlando.

Singh, A., Sharma, A. & Dubey, G. (2020). Big data analytics predicting real estate prices. *International Journal of System Assurance Engineering and Management*, 11(S2), 208–219.

Tchuente, D. & Nyawa, S. (2021). Real estate price estimation in French cities using geocoding and machine learning. *Annals of Operations Research*, 308, 571-608.

2 From Human Business to Machine Learning – Methods for Automating Real Estate Appraisals and their Practical Implications

2.1 Abstract

Until recently, in most countries, the use of Automated Valuation Models (AVMs) in the lending process was only allowed for support purposes, and not as the sole value-determining tool. However, this is currently changing, and regulators around the world are actively discussing the approval of AVMs. But the discussion is generally limited to AVMs that are based on already established methods such as an automation of the traditional sales comparison approach or linear regressions. Modern machine learning approaches are almost completely excluded from the debate. Accordingly, this study contributes to the discussion on why AVMs based on machine learning approaches should also be considered. For this purpose, an automation of the sales comparison method by using filters and similarity functions, two hedonic price functions – namely an OLS model and a GAM model, as well as a XGBoost machine learning approach, are applied to a dataset of 1.2 million residential properties across Germany. We find that the machine learning method XGBoost offers the overall best performance regarding the accuracy of estimations. Practical application shows that optimization of the established methods - OLS and GAM - is time-consuming and labor-intensive, and has significant disadvantages when being implemented on a national scale. In addition, our results show that different types of methods perform best in different regions and, thus, regulators should not only focus on one single method, but consider a multitude of them.

Keywords – Automated Valuation Models, Extreme Gradient Boosting, Housing Market, Machine Learning, Sales Comparison Method

2.2 Introduction

Although the real estate industry is often accused of lagging behind in terms of digitalization, the automation of processes is in fact being more and more actively discussed. In addition to the potential cost savings, ongoing improvements of computer resources and available data play an important role. Hence, it is now possible to raise data potential by automating daily processes. This potential can be leveraged in all areas of the real estate industry. Focusing on valuation, Automated Valuation Models (AVMs) have the power to change the appraisal process in many ways.

In the real estate industry, there are three different approaches to assessing properties, namely the cost approach, the income approach, and the sales comparison approach (see, e.g., Schulz et al., 2014 and Kim et al., 2020). The latter assumes that the value of a property can be derived from the value of comparable properties, and is particularly well suited for automated real estate valuations. Various ways are known in science and practice to apply the sales comparison approach in the context of AVMs (see, Isakson, 2002). Besides the integration of filters and similarity functions, well-established hedonic price models and modern machine learning approaches can also be used for AVM construction (see, e.g., Pagourtzi et al., 2003 and Bogin & Shui, 2020). Furthermore, repeated sales methods are employed for AVM applications, see, e.g., Oust et al. (2020).

Currently, the use of AVMs in the lending process is only allowed for supporting purposes in most countries and not as a value-determining tool (Matysiak, 2017 and Downie & Robson, 2008). Although there are now regulatory efforts to include AVMs in the lending process, this is only possible if the traceability, auditability, robustness and resilience of the inputs and outputs can be guaranteed (European Banking Authority, 2020). However, it remains unclear which of the abovementioned methods meet these requirements. While there is an ongoing debate about allowing the use of AVMs based on already established methods such as similarity functions or OLS regressions within the lending process, the application of modern machine learning methods is almost completely absent from the regulatory discussion. This is in fact due to the “black box” label of modern machine learning techniques. The decisions made by these methods are not as easy to understand as is the case for linear-based models due to more complex internal processes. However, in recent years, there have been various approaches to opening this black box; see for example by Friedman (2001), Goldstein et al. (2015), Lundberg & Lee (2017) and Apley & Zhu (2020). Through these approaches, the requirements of the supervisory authority for tractability and audibility can be considered.

Therefore, the question arises as to whether modern machine learning algorithms should also be considered by the regulatory body. The objective of this paper is to contribute to this ongoing debate and deliver further insights, based on a unique nationwide dataset, into the optimal use of modern machine learning algorithms for AVMs from a theoretical and practical point of view. For this purpose, an automation of the sales comparison method by using filters and similarity functions, referred to as Expert Function (EXF), two hedonic price functions based on Ordinary Least Squares (OLS) and Generalized Additive Models (GAM), as well as the machine learning approach eXtreme Gradient Boost (XGBoost), are compared with each other.

We are the first to use a unique dataset of around 1.2 million market values of standard residential properties across Germany between 2014 and 2020, provided by a large German Banking Group, to test the four selected AVM approaches with respect to the question of whether the application of modern machine learning algorithms on a nationwide level is superior to the other approaches. The market values are based on appraiser valuations and can thus be assumed to be objective property values – unlike, for example, listing data.

The German real estate market is characterized by many different local markets whose development is often mutually independent. While metropolitan regions have seen a significant rise in values in recent years, property values in rural areas have stagnated in some cases. We are therefore also interested in whether there is one type of model which performs best in varying submarkets or whether there are structural differences. Due to the low population density, fewer observations are available in rural areas, which also raises the question of whether data availability has an impact on model performance and whether this has an influence on the decision to use machine learning algorithms for AVMs or not.

Hence, we contribute to the literature by addressing the following three research questions:

I. Do machine learning methods outperform well-established AVM methods like the OLS, the GAM and the EXF, and should they therefore also be considered within the regulatory discussion of AVMs?
II. Should AVMs rely on the use of one single approach, or should multiple models be integrated for different spatial areas?
III. Does the performance of the methods depend on data availability and structure?

Although AVMs represent a wide field in the literature, we are - to the best of our knowledge – the first to compare a filter- and similarity-based AVM approach, two well-established hedonic methods and a modern machine learning approach on a nation-wide level. Our results provide important insights into the practical application of AVMs and the discussion as to whether the usage of machine learning algorithms for the lending process should be allowed from a regulatory perspective.

We find that the machine learning method XGBoost offers the best performance regarding estimation accuracy. The EXF provides the highest transparency, but lower accuracy, as it tends to underestimate and does not allow calculation of the influences of individual property characteristics. The OLS and GAM are capable of doing so, but are most often outperformed by the XGBoost. Another advantage of the XGBoost is its high flexibility. While the optimization of the OLS and the GAM must be mainly done manually to achieve good model performance, the XGBoost automatically detects relevant patterns in the data. Therefore, this algorithm is better suited in practice to performing estimations based on large and complex datasets, such as nation-wide real estate valuations. However, our results also show that it is not advisable to focus on only one method when designing a nation-wide AVM. Although the XGBoost performs best across Germany, there are also regions where the EXF, the OLS or the GAM perform best. In this respect, the data availability within regions plays an important role and it is apparent that the strength of the machine learning approach cannot be improved in regions with limited training data. We therefore generally recommend testing several algorithms per region before making a final choice. In summary, our study shows that the use of machine learning algorithms for AVMs is beneficial in many situations and therefore, their approval should indeed be discussed by the regulatory authorities.

2.3 Literature Review

The following section provides a general overview of the existing literature in the field of AVMs. Due to the generally high attention devoted to this topic by the scientific community, numerous publications can be found dealing with AVMs.

The sales comparison approach normally uses a limited set of similar properties to evaluate the market value of a property, as described by French & Gabrielli (2018). Since the beginning of the computer assisted mass appraisal (CAMA) era, this approach has been automated by various researchers and is widely used in practice, especially in North America and the UK. Usually, the designed approaches follow a predefined process to identify the n most comparable sales properties from a set of N observations. The final estimation is then calculated by taking the mean or similarity-weighted mean of these comparable sales prices. Early adoptions of the similarity-based finding of comparable properties can be found in Underwood & Moesch (1982), Thompson & Gordon (1987), Cannaday (1989), McCluskey & Anand (1999) and Todora & Whiterell (2002). More recently, Brunauer et al. (2017) design an approach for valuations of self-used property based on the sales comparison method. Trawinski et al. (2017) examine the accuracy of two expert algorithms, using either the N-Latest Transactions in an area (LTA) or the N-Nearest Similar Properties (NSP), and compare their results with different data-driven regression models. Ciuna et al. (2017) create an approach to overcome the limitations of AVMs in markets with less available data, by means of measuring the similarity degree of the comparables. Kim et al. (2020) automate the sales comparison method to evaluate apartments in Korea and find that their approach outperforms machine learning methods. Larraz et al. (2021) use a computer-assisted expert algorithm and consider differences in characteristics compared to similar properties and their relative location.

As Borst & McCluskey (2007) show, the similarity-based automation of the sales comparison approach is also reflected in spatial autoregressive (SAR) models. The authors state that the automated sales comparison approach can be seen as a special case of a spatially lagged weight matrix model, and that there is also a less formal but clear relationship with geographically weighted regressions (GWR). Applications of SAR models can be found, among other, in McCluskey et al. (2013) and Schulz & Wersing (2021). Compared to the approach of similarity-based finding of comparable properties, the SAR model is a much more complex approach and is associated with a higher computing cost.

The hedonic price function is a well-established model that has been widely used in research for decades and was primary described by Rosen (1974). Hedonic price models do not start from the property to be valued, but from the existing information on any property available in the market, as outlined by Maier & Herath (2015). Accordingly, the property value comprises an aggregation of various attributes or characteristics regarding the amenities, micro/macro location and geodata. This also allows conclusions to be drawn about the influence of individual attributes on the value. Based on Ordinary Least Square Regression (OLS), various studies use this method in real estate valuation, for example Malpezzi (2003), Sirmans et al. (2005) and Schulz et al. (2014). In the most recent studies, OLS is used as a benchmark, for example by Zurada et al. (2011), Chrostek & Kopczewska (2013), Cajias et al. (2019) and Chin et al. (2020). For the interested reader, Metzner & Kindt (2018) and Mayer et al. (2019) provide a detailed literature review of OLS in real estate valuation.

One main disadvantage of the OLS is the dependence on the correctly specified form of the independent variables, as described by Mason & Quigley (1996). As an advanced regression model, the GAM can overcome this drawback, as it can model non-linear relationships. So-called splines are used to non-parametrically describe the relationship between the dependent and independent variables. The model was first introduced by Hastie & Tibshirani (1990) and is based on the Generalized Linear Model established by Nelder & Wedderburn (1972). Investigating the housing market in Los Angeles, Mason & Quigley (1996) are the first to use a GAM in a real estate context and find statistically significant advantages compared to OLS models. The greater flexibility and increased accuracy enable GAMs to gain further acceptance in real estate price estimation. Various other studies deal with the application of GAMs for real estate valuation, namely Pace (1998), Bao & Wan (2004), Bourassa et al. (2007), Bourassa et al. (2010) and Brunauer et al. (2010). For a detailed literature review, see Cajias & Ertl (2018).

Improved data availability and computational power have led to a whole new wave of machine learning methods, and their application to AVMs has become a widely discussed topic within academia. Machine learning methods are designed to identify non-linear structures. In addition to Artificial Neural Networks (ANN) and Support Vector Machines (SVM), tree-based models are most applied in the context of AVMs.

The idea of tree-based models dates back to Morgan & Sonquist (1963) and their automatic interaction detection (AID). The first decision tree algorithm was introduced by Quinlan (1979). The currently most commonly cited and used algorithm for decision trees was introduced by Breiman et al. (1984). Single decision trees are associated with the disadvantage that they easily overfit and therefore might perform worse on unseen data. To overcome this problem, ensemble learning techniques are used (Prajwala, 2015). Ensemble learning is defined as the combination of many "weak-learners" (e.g., single regression trees) to form one single "strong learner" (Sagi & Rokach, 2018). One efficient and commonly used version is the gradient boosting technique. The idea of gradient boosting originates back to Breiman (1997) and was primary introduced for regression trees by Friedman (2001). As Kok et al. (2017) describe, gradient-boosting models build many small decision trees subsequently, from residual-like measures of the previous trees and each tree is built from a random subsample of the dataset. Applied in real estate context, Ho et al. (2021) evaluate property prices in Hong Kong using gradient boosting trees and find that this approach outperforms other machine learning techniques like Support Vector Machines (SVM). Another example can be derived from Singh et al. (2020). The authors compare the result of gradient boosting machines with the results of a random forest regression and a linear regression approach for housing sale data in Ames, Iowa. Their findings confirm the superiority of the gradient boosting approach. Other examples can be found at Pace & Hayunga (2020) and Tchuente & Nyawa (2021). Based on the concept of gradient boosting, Tianqi & Guestrin (2016) implement the eXtreme Gradient Boosting (XGBoost) algorithm. The XGBoost is a computationally effective and highly efficient version of gradient boosting trees and applies a more regularized model structure, in order to control overfitting. Since its introduction it has often been used to tackle real-estate-specific problems. Kumkar et al. (2018), for example, compare four tree-based ensemble methods, namely bagging, random forest, gradient boosting and eXtreme gradient boosting, in terms of their efficiency in the appraisal of property in Mumbai, India. Their findings show that the XGBoost model performs better than to the other models. Sangani et al. (2017) compare the results of different gradient boosting specifications with a simple linear regression. Their analysis is based on a dataset of 2,985,217 parcels in three different counties of California. The XGBoost gradient boosting specification significantly outperforms the linear regression and is also able to perform better than almost all other specifications. Further applications of the XGBoost algorithm can be seen in Kok et al. (2017), Cajias et al. (2019) and Birkeland et al. (2021).

Although AVMs represents a wide field in the literature – to the best of our knowledge – there is currently no research comparing the performance of an advanced machine learning approach with both a filter- and similarity-based AVM and a well-established hedonic model on a nation-wide level. To address this gap in the literature, we design our own filter- and similarity-based AVM, named EXF, and apply two frequently used hedonic models, to compare their results against the performance of a modern machine learning algorithm. We use the XGBoost as our machine learning model. In several other studies, the XGBoost shows encouraging results and, compared to ANNs and SVMs, has the advantage that calculation is quicker and is therefore best suited for the size of our data set. For the hedonic models, we decide to use an OLS and a GAM. The OLS is considered to be the most widely used method in the field of AVMs and is commonly used as a benchmark. Therefore, its results are easy for readers to understand, interpret and classify. The GAM is a further development of the OLS, which can consider non-linearities by means of splines. The results of the GAM are therefore an important extension to those of the OLS. The GAM also demonstrates good performance in many other studies. Our comparison allows us to provide important insights with respect to the practical application of AVMs and the discussion on whether the usage of machine learning algorithms for the lending process should be allowed from a regulatory perspective or not.

2.4 Data

Our analysis is based on a data set of 1,212,546 residential properties across Germany. The data set is provided by a large German banking group and originates from valuations of standard residential real estate lending. The data was collected between 2014 and 2020. Table 2.1 shows how the observations are distributed over time. As the numbers show, there is a slight decreasing trend which is caused by market fluctuations. Especially, in 2020 due to COVID-19 restrictions, fewer valuations took place.

Table 2.1: Observations per year

	2014	2015	2016	2017	2018	2019	2020
n	196318	196403	176238	163365	165106	165996	149120
(%)	0.1619	0.1620	0.1453	0.1347	0.1362	0.1369	0.1230

All properties are georeferenced, making it possible to add a spatial gravity layer in order to account for spatial information. Features describing the location and neighborhood of the observations are added via Open Street Map and Acxiom[1]. The dataset was cleaned for possible outliers, erroneous values, and incompleteness.

The observations are distributed across Germany and categorized into 327 administrative districts. The division of these regions is aligned with the NUTS-3 nomenclature of the European Union. The exact distribution of individual observations can be seen on the left side of Figure 2.1. Most observations are located around the largest German metropolitan areas like Berlin, Hamburg and Munich. In addition, a difference can be observed between west and east Germany, with the east tending to have fewer observations. This is consistent with the widely diverging population figures between these regions. A comprehensive introduction to the structure of the German regions can be found at Just & Schaefer (2017), and a more detailed overview of the German real estate markets is given by Just & Maennig (2012).

The market value of the properties, based on professional appraiser valuations, is used as the target variable. In contrast to listing data, market values do not depend on subjective seller perceptions of value, but are assessed objectively by outside third parties. An overview of the average market values across the 327 administrative districts is provided on the right side of Figure 2.1. The areas with the highest market values can be found in

[1] Acxiom is an American provider of international macroeconomic and microeconomic data. Further information can be found at: https://www.acxiom.com/.

the so-called Top-7[2] cities and their commuter belts. Furthermore, the market values are by far the highest in the south of Germany and tend to be lower in the east.

Figure 2.1: Number of observations and average market value per district

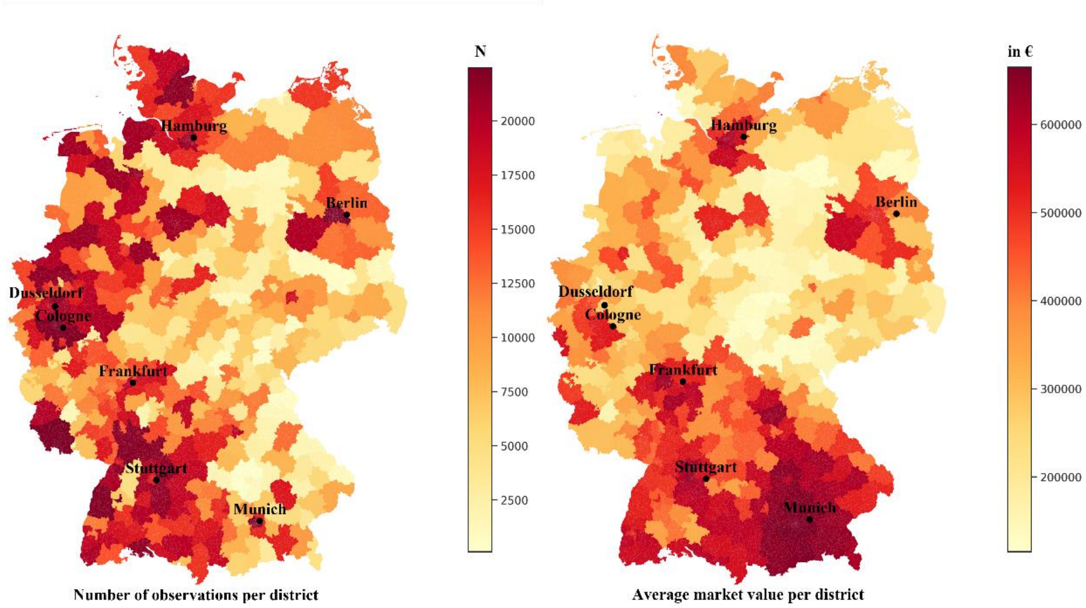


Table 2.2 shows the features included in our models and summarizes their univariate distributions. In principle, features describing the structural characteristics, micro-location and macro-location of the properties are selected. In addition, the year and quarter of the valuation is used to capture a temporal trend and seasonality. There are no correlations of concern within the data set, so that all variables can be integrated accordingly.[3]

The subtype of a property can be either a "Condominium", "Detached house" or "Not a detached house". The year of modernization represents the year in which the last major refurbishment took place. The use of the building describes the possible uses, either "Owner-occupied & Non-owner-occupied"[4], "Owner-Occupied" or "Non-owner-occupied". Basically, the variable describes whether a property can be rented to a third-party or not. The quality of the property is measured via a grade, on a scale ranging from 1 (very poor) to 5 (very good). The general condition of the property is represented by a categorial variable with 5 different categories ranging from disastrous to very good. The variable "Basement condominium" measures whether an apartment has an extra cellar compartment or not, whereas the "Basement" and "No Basement" variables are only valid for detached and non-detached houses. Features representing the micro-location and

[2] Berlin, Munich, Hamburg, Frankfurt am Main, Cologne, Dusseldorf, Stuttgart.
[3] The correlation matrix is available on request.
[4] Applies if the property is both partly owner-occupied and partly non-owner-occupied (e.g., single-family home with attached rental unit).

macro-location are latitude and longitude, different regiotypes, micro score and macro score of a location.

Table 2.2: Descriptive statistics

Variable	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Market value	Integer	228157.1	200000.0	141717.5	3860000.0	20100.00
Modernization year	Integer	1989.10	1988.00	17.19	2020.00	1950.00
Construction year	Integer	1978.48	1981.00	29.77	2020.00	1900.00
Year of valuation	Integer	2016.82	2017.00	2.00	2020.00	2014.00
Quarter of valuation	Integer	2.45	2.00	1.12	4.00	1.00
Quality grade	Integer	3.12	3.00	0.51	5.00	1.00
Macro score	Float	47.61	47.03	11.20	86.50	9.77
Micro score	Float	72.73	74.20	14.44	99.85	0.00
Living area	Float	120.31	114.68	51.69	440.00	15.00
Lot size	Float	436.48	323.00	541.66	10000.00	0.00
Latitude	Float	50.62	50.74	1.85	55.02	47.40
Longitude	Float	9.25	8.94	1.90	19.25	5.87
Basement condominium	Binary	0.38	0.00	0.48	1.00	0.00
No basement	Binary	0.19	0.00	0.39	1.00	0.00
Basement	Binary	0.44	0.00	0.50	1.00	0.00
Owner-occupied & Non-owner-occupied	Binary	0.09	0.00	0.29	1.00	0.00
Owner-occupied	Binary	0.70	1.00	0.46	1.00	0.00
Non-owner-occupied	Binary	0.21	0.00	0.41	1.00	0.00
Object subtype condominium	Binary	0.38	0.00	0.48	1.00	0.00
Object subtype detached house	Binary	0.42	0.00	0.49	1.00	0.00
Object subtype not a detached house	Binary	0.20	0.00	0.40	1.00	0.00
Condition good	Binary	0.38	0.00	0.49	1.00	0.00
Condition disastrous	Binary	0.00	0.00	0.02	1.00	0.00
Condition middle	Binary	0.45	0.00	0.50	1.00	0.00
Condition moderate	Binary	0.02	0.00	0.14	1.00	0.00
Condition bad	Binary	0.00	0.00	0.05	1.00	0.00
Condition very good	Binary	0.15	0.00	0.36	1.00	0.00
Regiotype agglo commuter belt	Binary	0.15	0.00	0.36	1.00	0.00
Regiotype agglo CBD	Binary	0.13	0.00	0.34	1.00	0.00
Regiotype agglo middle order centre	Binary	0.13	0.00	0.34	1.00	0.00
Regiotype agglo upper order centre	Binary	0.04	0.00	0.19	1.00	0.00
Regiotype rural commuter belt	Binary	0.15	0.00	0.36	1.00	0.00
Regiotype rural middle order centre	Binary	0.07	0.00	0.26	1.00	0.00
Regiotype rural upper order centre	Binary	0.01	0.00	0.07	1.00	0.00
Regiotype urban commuter belt	Binary	0.15	0.00	0.36	1.00	0.00
Regiotype urban middle order centre	Binary	0.10	0.00	0.29	1.00	0.00
Regiotype urban upper order centre	Binary	0.07	0.00	0.26	1.00	0.00

Notes: The parameter “market value” is the dependent variable in the model estimation.

The regiotype was provided by Acxiom, and clusters Germany into ten different area types. In general, Acxiom defines four different spatial types: “Central-Business-District”, "Agglomeration Area", "Urban Area" and "Rural Area". The last three can be divided further into three sub-categories each ("Upper Centers", "Middle Centers", "Commuter Belt"). All addresses in Germany can be allocated to one of the ten area types. The individual area types are determined according to the respective settlement structure and population density within the municipality and its surrounding area. In most cases, the selected NUTS-3 regions can be divided further into different Regiotypes and therefore, the integration of different subtypes enables taking further local fixed effects into account.

The micro score of a location is calculated via a gravity model and reflects accessibility in the sense of proximity to selected everyday destinations. A gravity model is a common method for approximating the accessibility of a location and is based on the assumption that nearby destinations play a greater role in everyday life than more distant ones (Handy and Clifton (2001)). The score is mainly used to reduce dimensionality and complexity for the EXF. The relevant points-of-interest (POIs) are selected from the findings of Powe et al. (1995), Metzner & Kindt (2018), Yang et al. (2018), Nobis & Kuhnimhof (2018) and Huang & Dall'erba (2021) and are provided in Table 2.3. A more detailed description of the construction of the micro score of a location can be found in Appendix I.

Table 2.3: Features of the micro score of a location

Points-of-Interests	Category	Description
University	Education & Work	University campus: institute of higher education
School	Education & Work	Facility for education
Kindergarten	Education & Work	Facility for early childhood care
CBD	Education & Work	Center of the next city
Supermarket	Local Supply	Supermarket – a large store with groceries
Marketplace	Local Supply	A marketplace where goods are traded daily or weekly
Chemist	Local Supply	Shop focused on selling articles for personal hygiene, cosmetics, and household cleaning products
Bakery	Local Supply	Place for fresh bakery items
ATM	Local Supply	ATM or cash point
Hospital	Local Supply	Facility providing in-patient medical treatment
Doctors	Local Supply	Doctor's practice / surgery
Pharmacy	Local Supply	Shop where a pharmacist sells medications
Restaurant	Leisure & Food	Facility to go out to eat
Café	Leisure & Food	Place that offers casual meals and beverages
Park	Leisure & Food	A park, usually urban (municipal)
Fitness Centre	Leisure & Food	Fitness center, health club or gym
Movie Theater	Leisure & Food	Place where films are shown
Theater	Leisure & Food	Theatre where live performances take place
Shopping Mall	Leisure & Food	Shopping center – multiple stores under one roof
Department Store	Leisure & Food	Single large store selling a large variety of goods
Subway Station	Transportation	City passenger rail service
Tram Station	Transportation	City passenger rail service
Railway Station	Transportation	Railway passenger only station
Bus Stop	Transportation	Bus stops of local bus lines
E-Charging Station	Transportation	Charging facility for electric vehicles

Note: The descriptions of the selected Points-of-Interest is based on the explanations of Open Street Map.[5]

To account for further local fixed effects, a macro score of a location is computed. For calculation, we use a social area analysis introduced by Carpenter et al. (1955). The method assumes that a city or region can be divided into homogeneous sub-areas on the basis of different environmental variables. The variables used in our study can be seen in Table 2.4 and are available at ZIP code level. The feature selection is based on Metzner & Kindt (2018). Further information about the macro scores can be found in Appendix II.

Table 2.4: Features for the macro score of a location

Feature	Category	Description
Educational Level	Social Status	Household structure by educational qualifications
Unemployment Rate	Social Status	Proportion of unemployed
Proportion of Children	Social Status	Proportion of population under 6 years
Purchasing Power	Economic Status	Purchasing power per household
Income Structure	Economic Status	Household structure by income
Social Security	Economic Status	Proportion of employees with social security
Relocation Behavior	Real Estate Market	Difference between inflows and outflows
Population Forecast	Real Estate Market	Population forecast for the next 5 years
Building Permits	Real Estate Market	Proportion of building permits
Construction Completions	Real Estate Market	Proportion of construction completed
Time-on-Market	Real Estate Market	Time-on-Market of properties sold

2.5 Methodology

2.5.1 Expert Function

The EXF uses different filters and similarity functions to determine nearby and similar comparable properties. As a result, it provides a final list of m comparables, revealing the highest degree of similarity to the property being evaluated. The next step is to estimate the market value by taking the average value of these comparables. Overall, this approach replicates the practice of traditional real estate appraisers in an automated manner. Starting with a total of N observations, a filter for spatial proximity is applied first for the EXF. Only observations within a radius of 20 km from the property to be valued are considered. Second, objects are only selected if they have the same Acxiom regiotype. Third, another filter is used to eliminate observations whose valuation date is too far in the past (< 5 years).[6] Other filters are set for the object type, occupation and presence of a

[5] See https://wiki.openstreetmap.org/wiki/Map_features.

[6] For valuations longer than one year ago, an indexation with the Destatis Real Estate Price Index is applied. The index is available quarterly for five Destatis-Regiotypes starting in 2016. Mapping with the Acxiom Regiotype is performed. Further information about the index can be found at https://www-genesis.destatis.de/genesis/online.

basement, so as to select only corresponding observations. Finally, filters are set for condition and quality grade, eliminating any observations that deviate by more than one category.

After the filtering, $n \leq \mathrm{N}$ observations are left and compared with the object to be valued x^* with the aid of similarity functions. These are intended to reflect the appraiser's approach to the selection of similar properties and make it possible to select only the most similar observations for the final estimation of market value.

First, a function for spatial proximity $SP(x_i, x^*)$ is applied for all objects x_i, $i \in n$:

$$SP(x_i, x^*) = \begin{cases} 100 - 5 \cdot d(x_i, x^*), & \text{if } \; d(x_i, x^*) \; \in [0; 20], \\ 0, & \text{else,} \end{cases} \tag{1}$$

where $d(x_i, x^*)$ measures the distance between the objects as a network distance measure in kilometers (km). Next, a triangular function for measuring the similarity of the remaining features is applied:

$$tr(x_{i,f}, x_f^*, a) = \begin{cases} 100 - a(\left|x_{i,f} - x_f^*\right|), & \text{if } \left|x_{i,f} - x_f^*\right| < \dfrac{100}{a}, \\ 0, & \text{else,} \end{cases} \tag{2}$$

with $x_{i,f}$ being the value of feature f of observation i and x_f^*, the corresponding features of the object being evaluated. a describes the slope of the function. A set of different slopes was tested to find the best parameters, yielding a to be 10 for the following features: construction year, modernization year, micro score and macro score and 25 for living area and plot size.

For all objects n, we are now able to compute the feature-related similarities. These are used to calculate the overall similarity score between all x_i and x^*:

$$s(x_i, x^*) = SP(x_i, x^*) \cdot w_1 + \sum_{f=2}^{7} tr\left(x_{i,f}, x_f^*, a\right) \cdot w_f, \qquad i \in \{1, \ldots, n\}, \tag{3}$$

with $w_1 = \frac{1}{7}$ and $w_f = \frac{1}{7}$, for all $f \in \{2, \ldots, 7\}$.

Now, we have the similarity score of the finally filtered objects n. The next step is to find the m most similar objects to $x^*, m \leq n$. Therefore, we construct a new vector v, that includes the objects in a sorted manner, so that the object with the highest overall similarity score is in the first entry and the object with the lowest overall similarity score is in the last entry. Only the first m objects of v, and therefore m most similar objects, are considered to evaluate the estimated market value of x^* by averaging their market values:

$$f(x^*) = \frac{1}{m}\sum_{i=1}^{m} f(x_i). \tag{4}$$

In this paper, the five most similar objects are used to estimate the market value of x^*, which is the minimum number of comparables required by law to perform a valuation by the sales comparison approach in Germany.[7]

2.5.2 Ordinary Least Square Regression – OLS

The first hedonic method we use is an OLS. This approach is the most commonly applied hedonic model and often used as a benchmark. Due to its simple architecture, it is easy to understand and interpret. The aim of an OLS is to explain a dependent variable y_i with independent variables $x_{i,1}, \dots, x_{i,k}$ and an error term ε_i:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k} + \varepsilon_i, \tag{5}$$

for all observations $i = 1, \dots, n,$ with

$$\mu_i = E[y_i] = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}. \tag{6}$$

Thereby, the unknown parameters $\beta_1, \dots, \beta_k$ are estimated. The OLS assumes that the relationship between the dependent variable and independent variables is linear in parameters. Furthermore, the error terms ε_i are considered to be independent and to have a constant variance. A more detailed description can be found in Fahrmeir et al. (2013).

In order to compare the performance of the models in due course, various optimizations of the OLS are carried out. To achieve the best possible prediction power, several statistical instruments like variable transformations, interaction terms and backward stepwise regression are applied. In contrast to modern machine learning models, these optimizations must be performed manually. With 36 independent variables in the model, 630 pairwise interactions result, which must be calculated and considered for 327 different districts, summing to roughly 206,010 interactions overall. This number can easily go into the millions when higher order interactions are also taken into account. This can be seen as a drawback of the OLS models.

[7] This procedure is based on the German guidelines for determining the mortgage lending value, see §4 BelWertV.

2.5.3 Generalized Additive Model – GAM

The GAM is a further development of the OLS and mainly based on the concept behind the Generalized Linear Model. The relationship between the expected value of the dependent variable and the independent variables can be modelled using a monotonic link function g, like the logarithm or the identity function. In addition, the GAM has the advantage of being able to include unspecified, non-parametric smoothing functions s_j, $j \in \{1, \dots, k\}$, of covariates. Consequently, we obtain the model:

$$g(\mu_i) = \beta_o + s_1(x_{i,1}) + \cdots + s_k(x_{i,k}). \tag{7}$$

The main advantage of the GAM compared to the OLS is its flexibility to model non-linear relationships. For the interested reader, we refer to Wood (2017).

Again, to account for locational differences, a combination of different statistical instruments like interaction terms and this time, additionally, different penalized spline types like cubic and thin plane splines have been used. Like the OLS, however, the GAM has the disadvantage that optimizations, such as the choice of spline function or interaction terms, must mainly be performed manually.

2.5.4 Extreme Gradient Boosting – XGBoost

Extreme Gradient Boosting is a tree-based ensemble learning method. The idea of ensemble learning algorithms is to combine many so-called weak learners h_m, in our case, single decision trees, into one strong learner h:

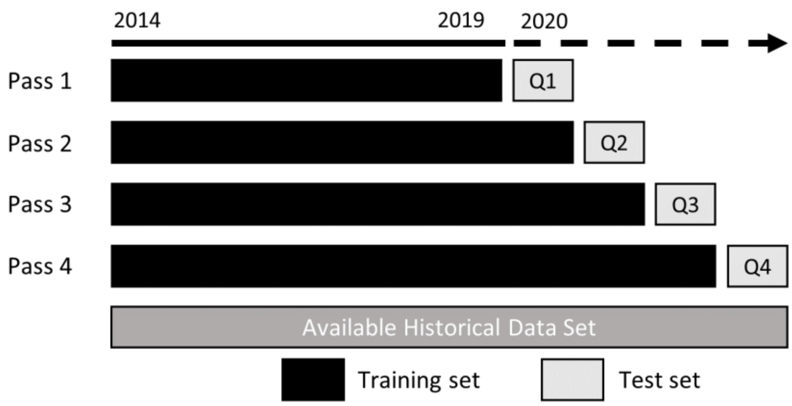
$$h(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{M} u_m h_m(\boldsymbol{y}|\boldsymbol{x}), \tag{8}$$

where u_m is used to weight the weak learners. M is the number of single trees, $\boldsymbol{x}$ is the full features space and $\boldsymbol{y}$ the response variable. Boosting is a type of ensemble learning in which the weak learners h_m are trained sequentially. Starting with one tree, the subsequent models learn from the previous errors. Gradient boosting uses the so-called gradient descent algorithm by adding new trees to minimize the loss of the model. The eXtreme Gradient Boosting is a computationally effective and highly efficient version of Gradient Boosting. In comparison to parametric and semi-parametric models, the XGBoost detects automatically complex patterns like non-linearities or higher-order interaction terms within a large amount of data, requiring for less manual optimization to account for location differences compared to the OLS and GAM. For more information about tree-based methods, ensemble learning and gradient boosting, the interested reader is recommended to read Hastie et al. (2001).

2.5.5 Testing Concept

To evaluate the predictive performance of the models, an extending window approach is implemented according to Mayer et al. (2019). Figure 2.2 illustrates the testing concept.

Figure 2.2: Extending window approach



The first iteration divides the dataset into a training set with observations from Q1/2014 to Q4/2019 and a test set from Q1/2020. In the next steps, the newly available data is added to the training set, and the models are retrained and tested on data of the next quarter. The advantages of this approach are that all algorithms are tested on unseen data and thus produce unbiased, robust results. Furthermore, the testing approach provides a realistic testing scenario. In Table 2.5, the number of training and test observations for each iteration are presented.

Table 2.5: Training and test observations

Data split	Q1	Q2	Q3	Q4
Training	1,063,426	1,106,866	1,141,612	1,180,741
Test	43,440	34,746	39,129	31,805

2.5.6 Evaluation Metrics

For each model, we compute the Mean Absolute Percentage Error (MAPE) and the Median Absolute Percentage Error (MdAPE) as accuracy measures. Unlike Mayer et al. (2019), we use the relative rather than the absolute measures of error to enable a more accurate comparison between administrative districts. Compared to the absolute measures, the relative measures provide a statement that represents the economic loss caused by the application of the algorithms much more precisely, which is very useful in our case, as we conduct a nationwide analysis involving many areas with varying levels of property market values. As Rossini & Kershaw (2008) and Ecker et al. (2020) state, the MAPE and MdAPE are two precision metrics, which enable a useful comparison across different models, datasets and locations. Other examples of their use can be found, for example, at Peterson & Flanagan (2009), Zurada et al. (2011), McCluskey et al. (2013), Schulz et al. (2014) and Oust et al. (2020).

At this point it should be mentioned that the economic loss for mortgage lenders is not symmetric as overvaluations in particular play a more critical role than undervaluations. Overvaluations significantly increase the potential risk that the value of a property does not cover a mortgage default (see, e.g., Shiller & Weiss, 1999). Both the MAPE and the MdAPE are not able to detect if there is a bias in a certain direction. To cover this topic, we additionally analyze a density plot of the relative deviations of the market values to the predicted values to investigate whether there is a bias in a certain direction or not.

In order to obtain an overall picture of the strengths and weaknesses of the algorithms, we additionally provide the proportion of predictions within 10 and 20 percent (PE(x)), as well as the coefficient of determination R^2. The ratio of error buckets (PE(x)) allows us to interpret the results in a simple and intuitive way for the human brain. They show how many of the observations can be estimated within a relative deviation of 10 or 20 percent. Schulz & Wersing (2021) state that the error buckets are frequently used by practitioners when assessing valuation accuracy. A detailed description of all metrics can be found in Table 2.6.

Table 2.6: Evaluation metrics

Error	Formula	Description
Mean Absolute Percentage Error (MAPE)	$MAPE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert$	Mean of all absolute percentage errors. A lower MAPE signals higher overall prediction accuracy in percent.
Median Absolute Percentage Error (MdAPE)	$MdAPE(y,\hat{y}) = median(\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert)$	Median of all absolute percentage errors. A lower MdAPE denotes a higher precision in percent without being sensitive to outliers.
Error buckets (PE(x))	$PE(x) = 100\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert < x$	Percentage of predictions where the relative deviation is less than x%, with x being 10 and 20. A larger PE(x) signals a lower variation in the predictions.
R^2	$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	Coefficient of determination. A high R^2 is an indication of better goodness of fit of the model.

2.6 Results

2.6.1 Results at National Level for Germany

Firstly, the models are compared at a national level. In Table 2.7, the prediction errors of the entire year 2020 are summarized. For all methods, the results of the metrics evolve similarly. The more complex the structure of the approach, the better the performance. The EXF is designed to replicate the practice of traditional real estate appraisers in an automated manner and is therefore readily understandable. However, the approach provides the poorest results. Comparing these results with the performance of the OLS, often used as a baseline model, we can see a performance improvement. Relatively speaking, the MAPE of the OLS is around 18% lower and the MdAPE 19%. In addition, using an OLS results in 18% and 20% more predictions deviating less than 10 and 20 percent from their actual market value.

Analyzing the results of the GAM, we again see a boost in performance compared to the OLS. But this time the relative improvement is smaller. The MdAPE of the OLS is around 9% higher. In addition, the percentage of predictions with a relative deviation of less than 10 and 20 percent increased by 9% and 5% respectively. This might be caused by the ability of the GAM to model more complex non-linearities within the data, which is extremely difficult to manually reproduce within the OLS, and practically impossible to implement for 327 districts. This is especially so, since these manual adaptions have to be done in each of the four quarters.

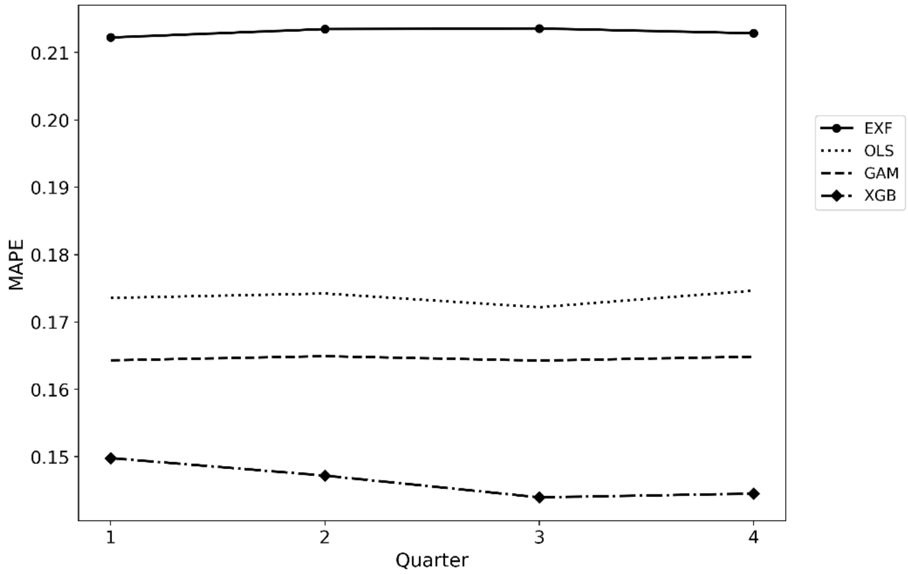
Overall, the XGBoost yields the best model performance regarding all evaluation metrics due to its ability to capture and process joint effects, non-linear relationships and high-dimensional structures within the data with comparably low manual effort. Comparing the results of the XGBoost with the EXF 43% and 33%, more observation deviate less than 10 and 20 percent from their market values.

Table 2.7: Model prediction errors 2020 throughout Germany

Models	MAPE	MdAPE	PE(10)	PE(20)	R^2
EXF	0.2130	0.1624	0.3267	0.5872	0.7735
OLS	0.1736	0.1311	0.3937	0.6940	0.8654
GAM	0.1646	0.1202	0.4273	0.7276	0.8664
XGB	0.1465	0.1084	0.4665	0.7786	0.8995

The chosen extending-window testing approach allows us to further analyze the performance of all four algorithms over the four quarters of 2020. Confirming the previous results, the solid line in Figure 2.3 shows the trends already mentioned. Additionally, it is interesting how consistently the models perform over all four quarters. Moreover, the XGBoost displays better performance the more training data it can process. The exact numbers can be seen in Appendix III.

Figure 2.3: MAPE plot

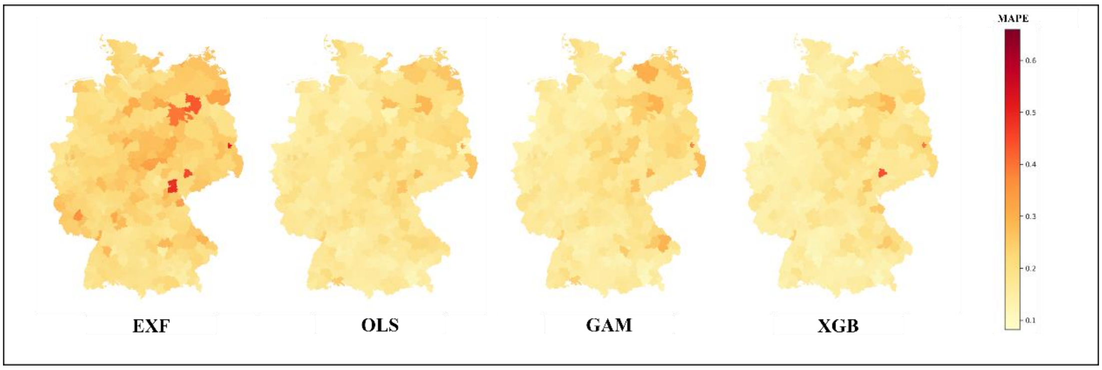


One research question of this study is to determine whether modern machine learning methods are able to outperform traditional hedonic models and the EXF approach. Analyzing our results at the national level for Germany we can clearly confirm this. The XGBoost yields a significant performance improvement compared to the EXF, OLS and GAM. This shows that in the future, regulators should also discuss the approval of machine learning methods in the field of AVMs. The application of machine learning approaches can lead to a reduction in the economic loss caused by the AVM. Machine learning algorithms are able to better assess possible risks within the lending process and can thus fulfill the actual purpose of a real estate valuation in a much more target-oriented manner.

2.6.2 Results at the Administrative District Level

After comparing the models at the national level, we want to examine the model performance in more detail. Therefore, we focus on the level of the 327 administrative districts. In Figure 2.4, the performance based on the MAPE for the different methods is shown cartographically. The maps confirm the abovementioned trends. The EXF again yields the overall poorest performance and again, it can be seen that the more complex the approach, the better the results. In addition, all four models are unsatisfactory with respect to estimating the market value in the same administrative districts. This can also be confirmed by the correlation matrices shown in Appendix IV. Especially in the eastern part of Germany, the MAPE tends to be higher. This result might be caused by the lower data availability in these regions.

Figure 2.4: Error comparison at administrative district level



To obtain a better understanding of the model performance at the administrative district level, we focus on the box plots of the MAPE in Figure 2.5. Those confirm the trend displayed in Figure 2.4. The EXF again yields the overall poorest results. It delivers the largest interquartile range, the longest whiskers and contains the most outliers. The XGBoost has the lowest median MAPE of all four models, whereas it has only two extreme outliers. In contrast, the GAM and especially the OLS have a smaller range of outliers. These results indicate that the XGBoost does not always display the best model performance and therefore, different models should be used for each administrative district.

Figure 2.5: Box plots of MAPE at administrative district level

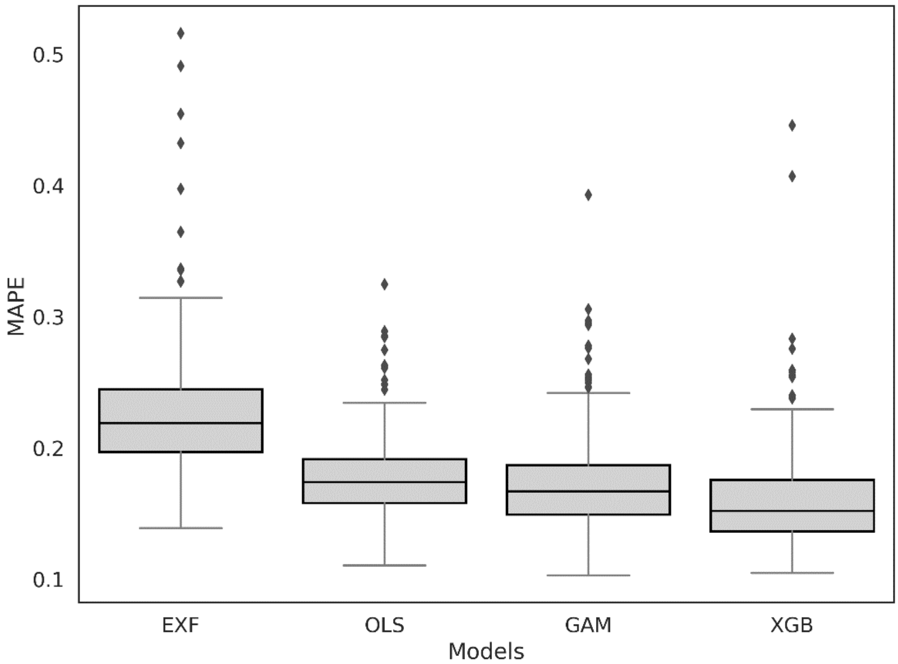


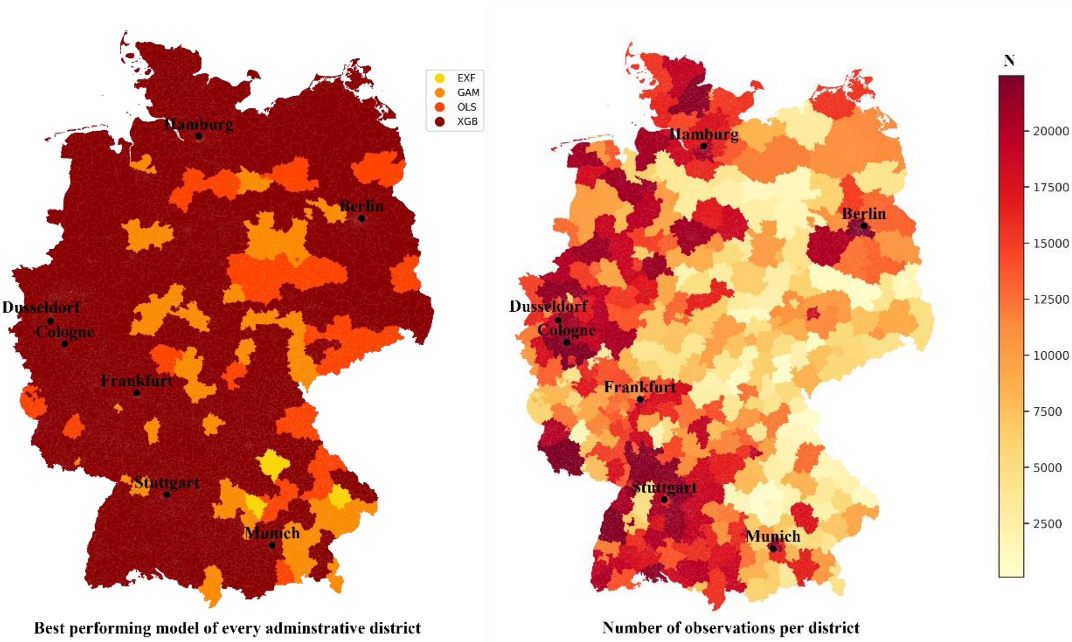
Table 2.8 shows the percentage of the administrative districts for which each model performs best. The XGBoost yields the best performance in all metrics for most administrative districts. Focusing on the hedonic approaches, the GAM and OLS are also superior in some regions, whereas EXF is the least convincing. The analysis shows that, in the case of Germany, there is no universally valid model that performs best in all administrative districts. Instead, it is advisable to apply different models in different regions.

Table 2.8: Model performance at administrative district level

Models	MAPE	MdAPE	PE(10)	PE(20)	R^2
XGB	0.7920	0.7187	0.6636	0.6636	0.6422
GAM	0.1162	0.1988	0.2202	0.2202	0.0550
OLS	0.0826	0.0765	0.1101	0.1101	0.2997
EXF	0.0092	0.0061	0.0061	0.0061	0.0031

To gain a deeper understanding of the finding that different models should be used in different regions, it is useful to present the results cartographically. On the left side of Figure 2.6, the best performing model regarding the MAPE in the administrative districts is shown. On the right, the number of observations per district is presented.

Figure 2.6: Model performance and number of observations per administrative district



In the north, west and south-west of Germany, the XGBoost shows the best model performance. In contrast, especially in the south-east and east, a different picture emerges. Comparing the availability of observations with these findings, a clear dependence can be derived. In areas with many observations, the XGBoost in particular can demonstrate its strengths. By contrast, in areas with only a few observations – mostly rural regions – the GAM and OLS can also convince. Consequently, especially if one aims to implement an AVM including several different locations with a different amount of data, multiple algorithms have to be considered. By testing different algorithms, the specifics of each region can be addressed, and thus, the best model for each region can be used. This ultimately leads to a reduction of the economic loss caused by the AVM. This result shows that regulators should generally consider approving of different algorithms, and that their focus should not be on only one type of procedure.

2.6.3 Results at the Prediction Level

Lastly, we analyze the relative deviations of the market values to the predicted values for all four models. In addition to the known evaluation metrics, with regard to the regulatory requirements, it is recommended to always perform an analysis at the prediction level to check whether overvaluations and undervaluations occur evenly, or whether the algorithms used exhibit a bias in one direction. In terms of choosing the right model from a practitioner's perspective, this can have a big impact and reduce financial risks from automated valuations in the long run. Accordingly, Figure 2.7 provides the density plots at the prediction level. It is evident that the EXF is negatively skewed, indicating that the approach underestimates market values to a greater extent. Transferring this point to practice shows that the use of the EXF may be more advantageous from a risk management perspective, since the economic loss caused by an incorrect estimate by the model is statistically lower. In the event of a loan default and a potential undervaluation by the EXF, the outstanding loan amount should more easily be recovered from the proceeds of a foreclosure sale than it would be the case if the property were overvalued. The curves of the OLS, GAM and the XGBoost are more symmetric and rather leptokurtic. This suggests that overvaluations and undervaluations occur more evenly, potentially increasing the risk of economic loss relative to the EXF.

Furthermore, a cumulative distribution function plot, shown in Figure 2.8, is used to reveal whether one method outperforms another stochastically. The XGBoost is superior to the other models, with the GAM and OLS in particular being very close. In contrast, a clear gap can be seen between the OLS and the EXF. This confirms the results from above, and shows again that it is important from the regulator side also to think about approving of machine learning methods in the area of AVMs.

Figure 2.7: Density plot of the relative deviation of the market values to the predicted values

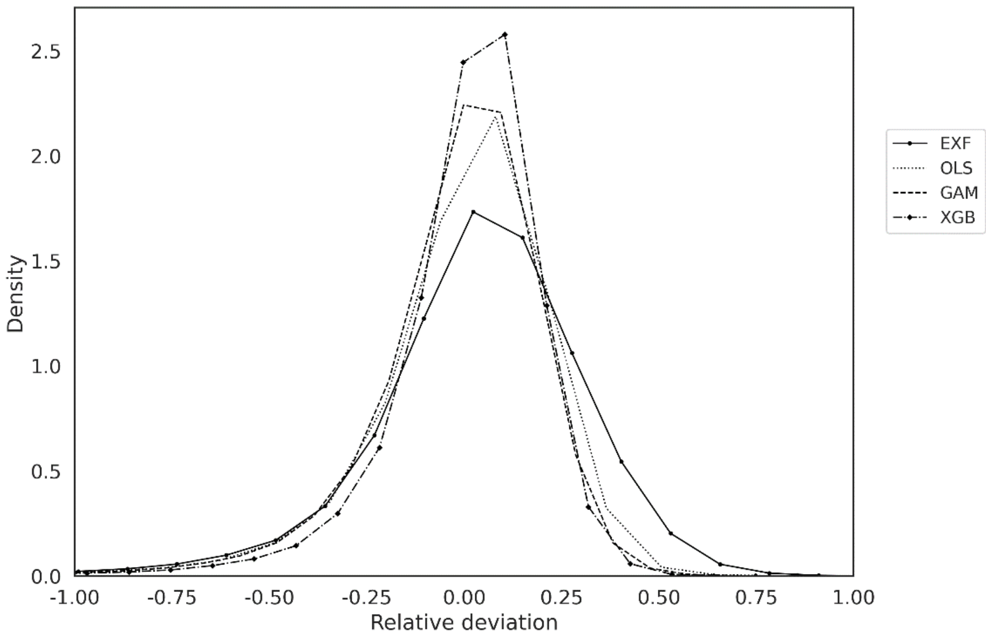
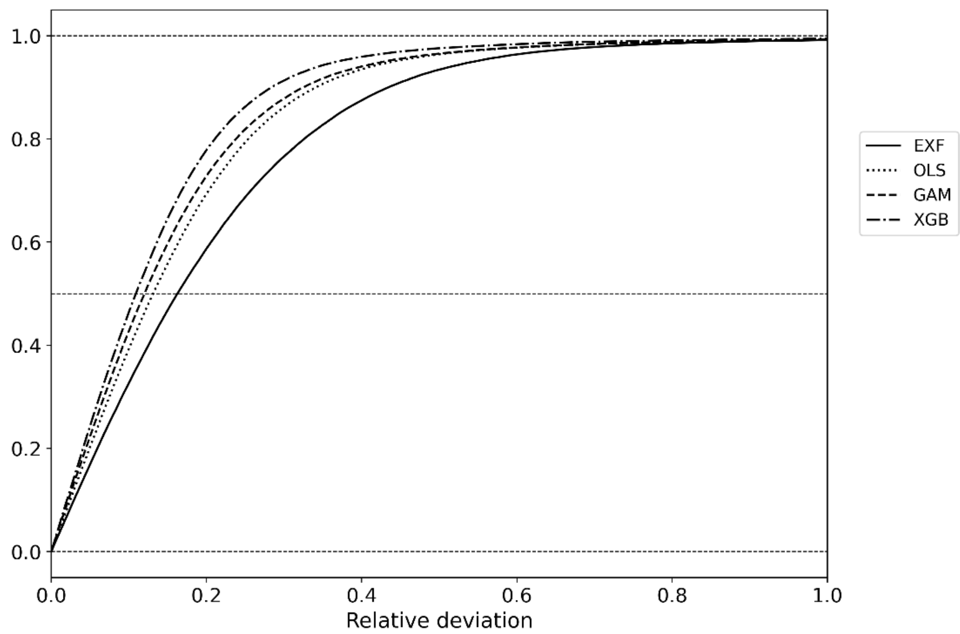


Figure 2.8: Cumulative distribution function plot of the relative deviation of the market values to the predicted values



2.7 Conclusion

This study compares different approaches to constructing AVMs on a nation-wide level in order to provide empirical evidence on the regulatory debate on the future use of automated valuations. In particular, we answer the question of whether more thought should also be given to the future use of machine learning algorithms in the context of AVMs. For this purpose, an automation of the sales comparison method by using filters and similarity functions - the EXF, two hedonic price functions based on OLS and GAM, as well as the machine learning approach XGBoost, are implemented for 327 administrative districts in Germany.

As our results show, the machine learning approach XGBoost achieves the highest overall accuracy (MAPE, MdAPE, PE(10), PE(20), R^2) in the valuation of standard residential properties in Germany. One reason might be its ability to automatically capture and process joint effects, non-linear relationships and high-dimensional structures within a large number of observations, without requiring as many manual optimizations to account for location differences. Therefore, the XGBoost convinces in practice with its flexibility. Especially in the metropolitan areas with many observations, the relationships between the variables determining the market value seem to be much more complex, implying a need for more complex valuation models. The OLS and GAM yield weaker results. Several optimizations have been carried out to increase their predictive performance and to ensure the comparability of the models as well. However, practical application shows that the optimization of the well-established methods is time-consuming, labor-intensive and in particular, therefore shows significant disadvantages in the implementation for 327 individual districts, as it is practically infeasible. Also, the EXF does not come close to the performance of the XGBoost. The EXF even shows the weakest performance compared to the XGBoost, the OLS and the GAM. Our results indicate that the EXF tends on average to underestimate the predicted market values.

Furthermore, the results of our study show that for designing an AVM, there is no “one size fits all”. Although the XGBoost is the best performer across the country, there are also administrative districts where the EXF, OLS, or GAM are best suited for estimating market values. In this context, it is particularly evident that the respective data availability seems to play a role. In districts with fewer observations, the traditional approaches manage to outperform the modern machine learning approach. In order to take this into account and to optimize the overall performance of AVMs, regulators should not merely allow, but actively promote the use of different types of algorithms. Before finally deploying an AVM, different types of methods should be tested for each district.

In the field of lending, a mispricing has major implications for both lenders and borrowers. Accurate model estimates are of considerable importance to ensure the resilience of the banking sector, especially in crisis periods. Our results clearly show that the approval of machine learning algorithms should be considered by regulators. We believe that machine learning algorithms have a high degree of robustness and resilience and are therefore ideally suited for AVMs. The traceability and auditability of the results required by the supervisory authorities can also be ensured by using the latest methods from the field of eXplainable Artificial Intelligence (XAI). While machine learning algorithms were considered as black box for a long time, XAI methods, like SHapely Additive exPlanations (SHAP) plots or Accumulated Local Effects (ALE) plots, are able to decode the basic decision-making process of any machine learning model. XAI is still at an early stage in the field of real estate research, but we are convinced that this will change in the coming years, and that new and important insights will be generated, which will further confirm the advantages of the use of machine learning algorithms. We therefore recommend re-examining the debate on the use of AVMs in everyday appraisals and, in particular, also including new and innovative methods.

2.8 Appendix

2.8.1 Appendix I – Micro Score

Our gravity model can be described using an activity function $f(A_p)$ and a distance function $f(D_{i,p})$:

$$A_{i,p} = \sum f(A_p) f(D_{i,p}). \tag{9}$$

$A_{i,p} \in [0,100]$ denotes the accessibility of point i for the POI p, whereby the activity function $f(A_p)$ specifies the relative importance of POI p, with $f(A_p) \in [0,1]$. $f(D_{i,p})$ measuring the travel time from point i to the POI p by using a non-symmetric sigmoidal distance function. The travel time was obtained for the selected POIs via Open Street Map and normalized using the following function:

$$L(x) = \frac{K}{(1 + Qe^{0.5x})^{\frac{1}{v}}}, \tag{10}$$

where $K, Q \in \mathbb{R}$ and $v \in \mathbb{R}^+$ are defined for all possible distances $x \in \mathbb{R}$. Furthermore, we have:

$$\begin{aligned} K &= (1+Q)^{1+v}, \\ Q &= v \cdot \exp(B \cdot x^*), \\ v &= \frac{\exp(B \cdot x^*) - 1}{\ln(y_i) - 1}, \end{aligned} \tag{11}$$

where x^* denotes a feature specific point of inflection and y^* is 0.5.

2.8.2 Appendix II – Macro Score

The scores $V_{j,i}(z)$ for each variable z in ZIP code i of region j are calculated using the following function:

$$V_{i,j}(z) = \left(\frac{100}{\max(z_j) - \min(z_j)}\right)\left(z_i - \min(z_j)\right), \tag{12}$$

where z_i denotes the value of feature z of ZIP code i. $\max(z_j)$, and $\min(z_j)$ are the maximum and minimum values of feature z in region j. As j, we define the 327 available administrative districts. Individual scores for all variables z included in the macro scores are calculated. The final macro score $MAS_{i,j}$ is computed by averaging the single scores in ZIP code i:

$$MAS_{i,j} = \frac{1}{|z|}\sum_{z} V_{i,j}(z). \tag{13}$$

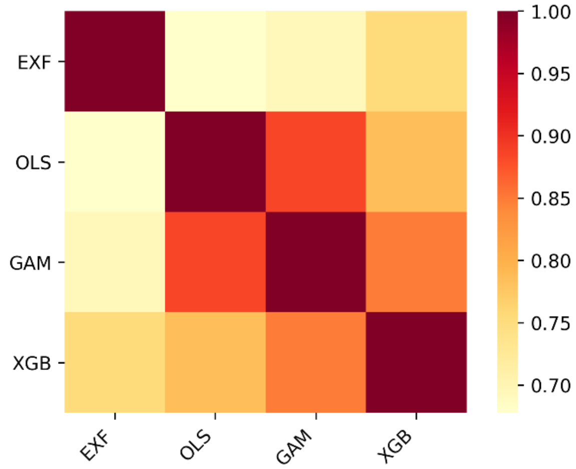
2.8.3 Appendix III – MAPE Results on a Quarterly Basis

Table 2.9: MAPE on a quarterly basis throughout Germany

Models	Q1	Q2	Q3	Q4
EXF	0.2122	0.2135	0.2136	0.2129
OLS	0.1736	0.1742	0.1722	0.1747
GAM	0.1643	0.1649	0.1643	0.1649
XGB	0.1498	0.1472	0.1440	0.1445

2.8.4 Appendix IV – District Error Correlation across the Models

Figure 2.9: District error correlation across the models



2.9 References

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.

Bao, H.X. & Wan, A.T. (2004). On the Use of Spline Smoothing in Estimating Hedonic Housing Price Models: Empirical Evidence Using Hong Kong Data. *Real Estate Economics*, 32(3), 487–507.

Birkeland, K.B., D'silva, A.D., Füss, R. & Oust, A. (2021). The Predictability of House Prices: "Human Against Machine". *International Real Estate Review*, 24(2), 139–183.

Bogin, A.N. & Shui, J. (2020). Appraisal Accuracy and Automated Valuation Models in Rural Areas. *The Journal of Real Estate Finance and Economics*, 60(1-2), 40–52.

Borst, R. & McCluskey, W. (2007). Comparative Evaluation of the Comparable Sales Method with Geostatistical Valuation Models. *Pacific Rim Property Research Journal*, 13(1), 106–129.

Bourassa, S.C., Cantoni, E. & Hoesli, M. (2007). Spatial Dependence, Housing Submarkets, and House Price Prediction. *The Journal of Real Estate Finance and Economics*, 35(2), 143–160.

Bourassa, S.C., Cantoni, E. & Hoesli, M. (2010). Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions. *Journal of Real Estate Research*, 32(2), 139–159.

Breiman, L. (1997). *Arcing the edge.* Technical Report 486, Statistics Department, University of California at Berkeley.

Breiman, L., Friedman, J., Stone, C.J. & Olshen, R.A. (1984). *Classification and regression trees.* CRC press.

Brunauer, W.A., Lang, S., Wechselberger, P. & Bienert, S. (2010). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *The Journal of Real Estate Finance and Economics*, 41(4), 390–411.

Brunauer, W.A., Weberndorfer, R.S. & Feilmayr, W. (2017). *A statistically founded sales comparison approach* [Conference Paper]. European Real Estate Society Meeting, Delft, Netherlands.

Cajias, M. & Ertl, S. (2018). Spatial Effects and Non-Linearity in Hedonic Modeling. *Journal of Property Investment & Finance*, 36(1), 32–49.

Cajias, M., Willwersch, J. & Lorenz, F. (2019). *I know where you will invest in the next year – Forecasting real estate investments with machine learning methods* [Conference Paper]. European Real Estate Society Meeting, Cergy-Pontoise Cedex, France.

Cannaday, R.E. (1989). How should you estimate and provide market support for adjustments in single family appraisals. *Real Estate Appraiser and Analyst*, 55(4), 43–54.

Carpenter, D.B., Shevky, E. & Bell, W. (1955). *Social Area Analysis: Theory, Illustrative Application and Computational Procedures.* Stanford University Press.

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. Rynek, Gospodarka, Społeczeństwo*, 35, 25–43.

Ciuna, M., Ruggiero, M. de, Manganelli, B., Salvo, F. & Simonotti, M. (2017). *Automated Valuation Methods in Atypical Real Estate Markets Using the Mono-parametric Approach.* In Computational science and its applications: 17th international conference, Trieste: Springer, 200–209.

Downie, M.-L. & Robson, G. (2008). *Automated Valuation Models: an International Perspective*. In RICS Automated Valuation Models Conference: AVMs Today and Tomorrow, London.

Ecker, M., Isakson, H. & Kennedy, L. (2020). An Exposition of AVM Performance Metrics. *Journal of Real Estate Practice and Education*, 22(1), 22–39.

European Banking Authority (2020). *Final Report: Guidelines on loan origination and monitoring.* European Banking Authority, Paris.

Fahrmeir, L., Kneib, T., Lang, S. & Marx, B. (2013). *Regression: Models, Methods and Applications*. Berlin: Springer.

French, N. & Gabrielli, L. (2018). Pricing to market. Property valuation revisited: the hierarchy of valuation approaches, methods and models. *Journal of Property Investment & Finance*, 36(4), 391–396.

Friedman, J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189–1232.

Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*, 24(1), 44–65.

Handy, S. L., & Clifton, K. J. (2001). Evaluating Neighborhood Accessibility: Possibilities and Practicalities. *Journal of Transportation and Statistics*, *4*(2-3), 67-78.

Hastie, T., Friedman, J. & Tibshirani, R. (2001). *The Elements of Statistical Learning.* New York: Springer.

Hastie, T. & Tibshirani, R. (1990). *Generalized additive models.* In *Statistical models in S,* New York: Routledge.

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.

Huang, Y. & Dall'erba, S. (2021). Does Proximity to School Still Matter Once Access to Your Preferred School Zone Has Already Been Secured?. *The Journal of Real Estate Finance and Economics*, 62(4), 548–577.

Isakson, H. (2002). The Linear Algebra of the Sales Comparison Approach. *Journal of Real Estate Research*, 24(2), 117–128.

Just, T. & Maennig, W. (Eds.) (2012). *Understanding German Real Estate Markets*. Berlin/Heidelberg: Springer.

Just, T. & Schaefer, P. (2017). *Germany's Regional Structure*. In Understanding German Real Estate Markets, Berlin/Heidelberg: Springer, 41–57.

Kim, Y., Choi, S. & Yi, M.Y. (2020). Applying Comparable Sales Method to the Automated Estimation of Real Estate Prices. *Sustainability*, 12(14), 5679.

Kok, N., Koponen, E.-L. & Martínez-Barbosa, C.A. (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal of Portfolio Management*, 43(6), 202–211.

Kumkar, P., Madan, I., Kale, A., Khanvilkar, O. & Khan, A. (2018). *Comparison of Ensemble Methods for Real Estate Appraisal*. In 2018 3rd International Conference on Inventive Computation Technologies (ICICT), IEEE.

Larraz, B., Alfaro-Navarro, J.-L., L Cano, E., Alfaro-Cortes, E., Garcia, N. & Gámez, M. (2021). A computer-assisted expert algorithm for real estate valuation in Spanish cities*. Environment and Planning B: Urban Analytics and City Science*, 48(6), 1712–1727.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Proc. *Advances in Neural Information Processing Systems*, 30, 4768–4777.

Maier, G. & Herath, S. (2015). *Immobilienbewertung mit hedonischen Preismodellen: Theoretische Grundlagen und praktische Anwendung*. Wiesbaden: Springer Gabler.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. *Housing Economics and Public Policy*, 1, 67–89.

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

Matysiak, G. (2017). *Automated Valuation Models (AVMs): A brave new world?*. In Wroclaw Conference in Finance 2017, Wroclaw.

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. & Anand, S. (1999). The application of intelligent hybrid techniques for the mass appraisal of residential properties. *Journal of Property Investment & Finance*, 17(3), 218–239.

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.

Metzner, S. & Kindt, A. (2018). Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties. *International Journal of Housing Markets and Analysis*, 11(1), 73–100.

Morgan, J.N. & Sonquist, J.A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American statistical association*, 58(302), 415–434.

Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370.

Nobis, C. & Kuhnimhof, T. (2018). *Mobilität in Deutschland – MiD: Ergebnisbericht*. Bundesministerium für Verkehr und digitale Infrastruktur, Bonn.

Oust, A., Hansen, S.N. & Pettrem, T.R. (2020). Combining Property Price Predictions from Repeat Sales and Spatially Enhanced Hedonic Regressions. *The Journal of Real Estate Finance and Economics*, 61(2), 183–207.

Pace, 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. *The Journal of Real Estate Finance and Economics*, 60(1-2), 170–180.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T. & French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, 21(4), 383–401.

Peterson, S. & Flanagan, A. (2009). Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research*, 31(2), 147–164.

Powe, N.A., Garrod, G.D. & Willis, K.G. (1995). Valuation of urban amenities using an hedonic price model. *Journal of Property Research*, 12(2), 137–147.

Prajwala, T.R. (2015). A Comparative Study on Decision Tree and Random Forest Using R Tool. *International journal of advanced research in computer and communication engineering*, 4(1), 196-199.

Quinlan, J.R. (1979). *Discovering rules by induction from large collections of examples*. In Expert systems in the micro electronic age, Edinburgh: University Press.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Rossini, P. & Kershaw, P. (2008). *Automated valuation model accuracy: Some empirical testing*. Doctoral dissertation, Pacific Rim Real Estate Society.

Sagi, O. & Rokach, L. (2018). Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4), e1249.

Sangani, D., Erickson, K. & Hasan, M.A. (2017). *Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting*. In IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Orlando.

Shiller, R.J. & Weiss, A.N. (1999). Evaluating Real Estate Valuation Systems. *The Journal of Real Estate Finance and Economics*, 18(2), 147–161.

Schulz, R. & Wersing, M. (2021). Automated Valuation Services: A case study for Aberdeen in Scotland". *Journal of Property Research*, 38(2), 154–172.

Schulz, R., Wersing, M. & Werwatz, A. (2014). Automated valuation modelling: a specification exercise. *Journal of Property Research*, 31(2), 131–153.

Singh, A., Sharma, A. & Dubey, G. (2020). Big data analytics predicting real estate prices. *International Journal of System Assurance Engineering and Management*, 11(S2), 208–219.

Sirmans, G.S., Macpherson, D.A. & Zietz, E.N. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*,13(1), 1–44.

Tchuente, D. & Nyawa, S. (2021). Real estate price estimation in French cities using geocoding and machine learning. *Annals of Operations Research*, 308, 571-608.

Thompson, J.F. & Gordon, J.F. (1987). Constrained regression modeling and the multiple regression analysis-comparable sales approach. *Property Tax Journal*, 6(4), 251–262.

Tianqi, C. & Guestrin, C. (2016*). XGBoost: A scalable tree boosting system*. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, New York: ACM, 785-794.

Todora, J. & Whiterell, D. (2002). Automating the sales comparison approach. *Assessment Journal*, 9(1), 25–33.

Trawinski, B., Telec, Z., Krasnoborski, J., Piwowarczyk, M., Talaga, M., Lasota, T. & Sawilow, E. (2017). *Comparison of expert algorithms with machine learning models for real estate appraisal*. In 2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA), IEEE.

Underwood, W.E. & Moesch, J.R. (1982). *The second generation of CAMA in New York State*. In First World Congress on Computer Assisted Valuation, Cambridge, MA.

Wood, S.N. (2017). *Generalized Additive Models: An Introduction with R*. Boca Raton: CRC Press.

Yang, J., Bao, Y., Zhang, Y., Li, X. & Ge, Q. (2018). Impact of Accessibility on Housing Prices in Dalian City of China Based on a Geographically Weighted Regression Model. *Chinese Geographical Science*, 28(3), 505–515.

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), 349–388.

3 Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values

3.1 Abstract

A sound understanding of real estate markets is of economic importance and not simple, as properties are a heterogenous asset and no two are alike. Traditionally, parametric or semi-parametric and, thus, assumption-based hedonic pricing models are used to analyze real estate market fundamentals. These models are characterized by the fact that they require a-priori assumptions regarding their functional form. Usually, the true functional form is unknown and characterized by non-linearities and joint effects, which are hard to fully capture. Therefore, their results should be interpreted with caution. Applying the state-of-the art non-parametric machine learning XGBoost algorithm, in combination with the model-agnostic Accumulated Local Effects Plots, (ALE) enables us to overcome this problem. Using a dataset of 81,166 residential properties for the seven largest German cities, we show how ALE plots enable us to analyze the value-determining effects of several structural, locational and socio-economic hedonic features. Our findings lead to a deeper representation of real estate market fundamentals.

Keywords: Housing Markets, Machine Learning, Explainable AI, Feature Importance, ALE Plots

3.2 Introduction

Understanding real estate markets and its drivers is arguably one of the most important areas of real estate research. Compared to other asset classes, real estate is a heterogeneous asset and properties differ from one another in terms of their features. There is a large body of literature dealing with the factors which have a significant influence on the value or price of a property, subsumed under the term Hedonic Price Models. They are usually based on parametric and semi-parametric methods like the Ordinary Least Square approach (see, e.g., Malpezzi, 2002; Sirmans et al., 2005; Schulz et al., 2014) or the Generalized Additive Models (see, e.g., Bourassa et al., 2007; Bourassa et al., 2010; Brunauer et al., 2010). In recent years, more advanced statistical and modern machine learning (ML) methods have attracted interest in the real estate community, as they are often less restrictive in terms of their model structure and thus are more flexible. Especially deep learning algorithms like Artificial Neural Networks (ANN), bagging techniques like random forest (RF) and boosting algorithms like the eXtreme Gradient Boosting (XGBoost) algorithm, seem better suited to real estate valuation.[8] In almost all cases, researchers conclude that machine learning techniques yield better predictions than standard linear models (see, e.g., Chun Lin & Mohan, 2011; Kok et al., 2017; Mayer et al., 2019). However, ML applications are usually criticized for their lack of transparency and are therefore often referred to as black boxes (see, e.g., Din et al., 2001; McCluskey et al., 2013). While parametric and semi-parametric applications are comprehensible to humans, the calculations of modern ML applications can only be understood with difficulty if at all. To overcome this problem, so-called eXplainable Artificial Intelligence (XAI) approaches have been developed. These approaches use model-agnostic frameworks to reveal the modes of operations of ML algorithms and thus help to make their mode of action more transparent.

In real estate, so far, XAI approaches have been explored only to a limited extent, but we believe they offer several benefits. First, they shed light on the mechanism underlying ML algorithms, thus overcoming their image of black boxes, and therefore increasing their acceptance in different regulated and unregulated areas within the real estate industry, for example in the mortgage lending industry. Second, XAI methods are able to support research in understanding the key drivers of real estate markets and their functional form

[8] Applications include Worzala et al. (1995), Din et al. (2001), Peterson & Flanagan (2009), McCluskey et al. (2013) and Chiarazzo et al. (2014) for neural networks. Antipov & Pokryshevskaya (2012), Bogin & Shui (2020) and Pace & Hayunga (2020) for random forests. Focusing on boosting-related methods, see van Wezel et al. (2005), Kagie & van Wezel (2007), Gu & Xu (2017), Sangani et al. (2017), Ho et al. (2021) and Stang et al. (2022).

by taking non-linearities and joint effects into consideration. These findings can among other things be used to validate or adapt the previously known understanding of econometric functional forms.

In order to further demonstrate and confirm these benefits empirically, we show in our study how XAI methods can be used to make the deep hidden patterns of residential real estate markets interpretable for human beings. Therefore, we use two different methods. First, we use Permutation Feature Importance (PFI), first introduced by Breiman (2001), to analyze which features actually influence the value of a property. Next, we use so-called Accumulated Effects Plots (ALE), established by Apley & Zhu (2020), to further make a statement about the effects themselves and whether non-linear relationships can be identified or not.[9] In particular, the former is used as a basis for the latter, to identify which variables have the greatest impact on property values. While PFI has been used – to a limited extent – for answering different real estate specific questions (see, e.g., Lorenz et al., 2022), we are – to the best of our knowledge – the first to use ALE plots to explore the determinants of residential property market values. Furthermore, the previous literature has mainly focused on the identification of non-linearities, but falls short on the interpretation of their economic implications. Accordingly, this paper not only focuses on identifying reliable and unbiased relations between features and residential property prices, but also discusses their economic implications.

To conduct our analysis, we use the modern ML algorithm eXtreme Gradient Boosting (XGBoost) and a unique dataset consisting of 81,166 residential properties for the seven largest cities of Germany. The dataset is from the years 2014 to 2020 and can be split into 61,763 condominiums and 19,403 single-family homes. We analyze the two groups separately in order to reveal differences between the two property subtypes, in addition to the general analysis of the value-determining features. We use the market value per square meter as our target variable. The underlying market values are based on appraiser valuations and are therefore verified by professional real estate appraisers.

Besides the general introduction of ALE plots in a real estate context, we contribute to the literature by addressing the following research questions:

I. Which characteristics are important for the market values of residential properties?

[9] In XAI research, Partial Dependence Plots (PDP) – proposed by Friedman (2001) – are one of the oldest and most widely used methods (see, e.g., Levantesi & Piscopo, 2020). However, PDP plots have been shown to produce biased results when features are correlated (Apley & Zhu, 2020). In real estate, many features have an intrinsic dependence that does not justify the use of PDP plots. In contrast, ALE plots do not have this disadvantage, and are therefore well suited to real estate market analysis.

II. To what extent are the features characterized by either linearity or non-linearity? Are there differences depending on different cities?

III. Are there fundamental differences between condominiums and single-family homes?

Our analyses reveal that the same value-determining features play an important role for both condominiums and single-family homes. However, there are fundamental differences within the two property types with regard to the shape of the individual ALE plots and thus the influence of the respective feature on the market value of a property. Furthermore, we identify non-linear relationships for the majority of features. Generalized rules of thumb such as "the larger the living area, the lower the market value per square meter" are refuted by our findings for condominiums, but can be confirmed for single-family homes. In summary, our results show how important it is for both real estate research and practice to conduct data-driven analyses with the help of modern ML and XAI approaches, in order to gain important market insights and, if necessary, to update long-established assumptions regarding the determinants of real estate market values. This is especially important for mortgage underwriters, valuation firms and regulatory authorities and, thus, of considerable interest to most of the real estate community.

3.3 Data

This study uses a dataset of 81,166 residential properties for the Top-7 cities of Germany. The data originate from the years 2014 to 2020. The Top-7 are the most important cities in Germany for the real estate industry and are: Berlin, Hamburg, Munich, Cologne, Frankfurt, Stuttgart and Dusseldorf. In comparison to other European countries such as England or France, the German real estate market is polycentric and not dominated by one large city. As Cajias & Freudenreich (2018) explain, analyzing the Top-7 cities leads to a "socially, culturally and economically well diversified overview of major urban areas all over Germany." As we are interested in analyzing differences between different subtypes of residential properties, the dataset is further split into two groups. The first consists of 61,763 condominiums and the second of 19,403 single-family homes. Table 3.1 shows how the individual observations are distributed among the seven cities.

The dataset is provided by a large German banking group and originates from their valuation department. The data was collected for traditional real estate lending to private customers and is used for the legally required valuation of the property serving as loan collateral. In principle, the banking group finances the entire range of properties available on the market across all cities. A potential selection bias (e.g., only financing a certain category of properties) can therefore be largely excluded. To avoid abuses, the valuations are not carried out by the lender itself, but by certified external third-party appraisers. The appraisals are carried out in accordance with the legal framework applicable in Germany and determined using the comparison approach.[10] The market value per square meter of the properties is used as the target variable. In addition to the dependent variable, a set of features defining the structural characteristics of the properties is used. All properties are georeferenced, making it possible to add a spatial gravity layer to account for spatial information. Features describing the location and neighborhood of the properties are added via Open Street Map and Acxiom.[11] The dataset is cleaned before being used to account for duplicates, incompleteness and erroneous data points.

Table 3.1: Observations per city and subtype

	Berlin	Hamburg	Munich	Cologne	Frankfurt	Stuttgart	Dusseldorf
Condominiums	15,166	5,703	12,743	13,189	5,559	4,108	5,295
Single-family Homes	6,545	3,555	1,408	4,933	1,140	1,008	814

In the area of structural characteristics, the construction year, living area, use of the property, condition and a quality grade were used for both apartments and single-family homes. Furthermore, the lot size, a variable describing whether the property has a basement or not and a feature outlining the subtype of the property, are used for the single-family homes. All these features were determined by professional appraisers in the context of their assessment process, which is why it can reasonably be assumed that these represent a detailed and truthful representation of the actual properties.

The use of the building describes the possible types of usage of the property, whereby the characteristics are either "Owner-occupied & Non-owner-occupied,"[12] "Owner-Occupied," or "Non-owner-occupied." Basically, the variable describes whether or not a property can be rented to a third-party. The quality of the property is measured in terms of a grade, on a scale ranging from 1 (very poor) to 5 (very good) and describes the general quality of the interior of a property (e.g., condition and quality of the bathrooms). The general condition of the property, on the other hand, refers to the exterior condition of the property (e.g., condition of the walls and facade) and is represented by a categorial variable with 5 different categories ranging from bad to very good. The features describing the subtype

[10] In the context of legally required real estate valuations in Germany, there are slight differences in the methodology used compared to the internationally common approaches. Detailed explanations can be found in Schnaidt & Sebastian (2012).

[11] Acxiom is an American data provider for international data. Further information can be found at: https://www.acxiom.com/.

[12] Applies if the property is both partly owner-occupied and partly non-owner-occupied (e.g., single-family home with attached rental unit).

of the single-family homes are binary, and state whether it's a detached or non-detached house.

Features representing the micro-location of a property are the latitude and longitude and four different micro scores. The micro scores of a location are calculated via a gravity model and reflect the accessibility as the proximity to selected everyday destinations for each category. A gravity model is a common method for approximating the accessibility of a location and is based on the assumption that nearby destinations play a greater role in everyday life than more distant ones (Handy & Clifton, 2001). The scores can range from 0 to 100 points, and the higher the score, the better the accessibility of the location. The relevant points-of-interest (POIs) are selected on the basis of the findings of Powe et al. (1995), Metzner & Kindt (2018), Yang et al. (2018), Nobis & Kuhnimhof (2018) and Huang & Dall'erba (2021) and are provided in Table 3.2. The scores are mainly used to reduce dimensionality and complexity. A more detailed description of the construction of the micro scores can be found in Appendix I.

Table 3.2: Features of the micro scores of a location

Points-of-Interests	Category	Description
University	Education & Work	University campus: an institute of higher education
School	Education & Work	Place for education
Kindergarten	Education & Work	Facility for early childhood care
CBD	Education & Work	Center of the next city
Supermarket	Local Supply	Supermarket – a large store with groceries
Marketplace	Local Supply	A marketplace where goods are traded daily or weekly
Chemist	Local Supply	Shop focused on selling articles of personal hygiene, cosmetics, and household cleaning products
Bakery	Local Supply	Place for fresh bakery goods
ATM	Local Supply	ATM or cash point
Hospital	Local Supply	Facility providing in-patient medical treatment
Doctors	Local Supply	Doctor's practice / surgery
Pharmacy	Local Supply	Shop where a pharmacist sells medications
Restaurant	Leisure & Food	Facility to go out to eat
Café	Leisure & Food	Place that offers casual meals and beverages
Park	Leisure & Food	A park, usually urban (municipal)
Fitness Centre	Leisure & Food	Fitness Centre, health club or gym
Movie Theater	Leisure & Food	Place where films are shown
Theater	Leisure & Food	Theatre or opera house where live performances occur
Shopping Mall	Leisure & Food	Shopping Centre– multiple stores under one roof
Department Store	Leisure & Food	Single large store selling a large variety of goods
Subway Station	Transportation	City passenger rail service
Tram Station	Transportation	City passenger rail service
Railway Station	Transportation	Railway passenger-only station.
Bus Stop	Transportation	Bus stops of the local bus lines.
E-Charging Station	Transportation	Charging facility for electric vehicles

Notes: The descriptions of the selected Points-of-Interests is based on the explanations of Open Street Map.[13]

[13] See https://wiki.openstreetmap.org/wiki/Map_features.

The macro-location is considered by means of the features "Unemployment Rate" and "Time-On-Market." These two features have been used frequently in other studies (see, e.g., Cheng & Fung, 2015). Both variables are available at ZIP Code level. The "Unemployment Rate" measures the percentage of workers in the labor force who do not currently have a job, but are actively looking for work, and is used as a proxy for the social status of the local inhabitants. The feature "Time-On-Market" is used as a proxy for liquidity and is defined as the average number of days properties are advertised on the market within a certain ZIP Code.

To capture a temporal trend and seasonality, the year and quarter of the valuation are included. There are no correlations of concern within the dataset, so that all variables can be integrated accordingly.[14]

Table 3.3 summarizes the descriptive statistics of the features used for condominiums and Table 3.4 those for single-family homes.[15]

Table 3.3: Condominium - Descriptive statistics

Variable	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Market value per square meter	Euro	3,691.20	3,254.55	1,911.97	1,8384.40	216.96
Living area	Sqm	72.19	69.00	28.34	203.57	15.0
Longitude	Coordinate	10.06	10.00	2.52	13.64	6.70
Latitude	Coordinate	50.80	50.95	1.79	53.68	48.07
Micro score - education and work	Percentage	94.59	97.88	7.58	99.89	0.00
Micro score - shopping	Percentage	88.84	92.78	11.04	99.29	0.00
Micro score - leisure	Percentage	98.84	99.64	3.35	99.98	0.00
Micro score - public transport	Percentage	64.22	67.86	19.22	97.90	0.00
Unemployment ratio	Percentage	6.27	5.60	4.42	26.89	0.04
Year of construction	Year	1974	1973	33.80	2020	1900
Year of valuation	Year	2016	2017	2.02	2020	2014
Quarter of valuation	Integer	2.45	2.00	1.12	4.00	1.00
Quality grade	Integer	3.19	3.00	0.52	5.0	1.00
Time on market	Weeks	10.50	9.90	4.23	60.7	2.80
Condition very good	Binary	0.18	0.00	0.38	1.00	0.00
Condition good	Binary	0.39	0.00	0.48	1.00	0.00
Condition middle	Binary	0.45	0.00	0.50	1.00	0.00
Condition moderate	Binary	0.01	0.00	0.11	1.00	0.00
Condition bad	Binary	0.00	0.00	0.04	1.00	0.00
Owner-occupied & Non-owner-occupied	Binary	0.10	0.00	0.31	1.00	0.00
Owner-occupied	Binary	0.44	0.00	0.50	1.00	0.00
Non-owner-occupied	Binary	0.46	0.00	0.50	1.00	0.00

[14] The correlation matrix is available on request.
[15] The individual summary statistics for each city are available on request.

Table 3.4: Single-family homes - Descriptive statistics

Variable	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Market value per square meter	Euro	3,064.06	2,693.19	1,538.35	2,2781.21	199.44
Living area	Sqm	133.68	126.43	42.19	402.00	73.07
Lot size	Sqm	467.92	396.00	296.63	1500.00	1.00
Longitude	Coordinate	10.25	10.03	2.64	13.75	6.70
Latitude	Coordinate	51.60	52.40	1.59	53.71	47.58
Micro score - education and work	Percentage	85.59	88.29	11.87	99.83	0.00
Micro score - shopping	Percentage	75.20	79.49	15.99	98.88	0.00
Micro score - leisure	Percentage	95.49	98.27	8.75	99.98	0.00
Micro score - public transport	Percentage	43.28	42.78	16.48	95.37	0.00
Unemployment ratio	Percentage	8.34	9.44	4.33	26.89	0.08
Year of construction	Year	1974	1977	30.18	2020	1900
Year of valuation	Year	2016	2016	1.97	2020	2014
Quarter of valuation	Integer	2.44	2.00	1.11	4.00	1.00
Quality grade	Integer	3.15	3.00	0.50	5.00	1.00
Time on market	Weeks	11.30	10.20	3.71	60.70	3.70
Condition very good	Binary	0.15	0.00	0.36	1.00	0.00
Condition good	Binary	0.42	0.00	0.49	1.00	0.00
Condition middle	Binary	0.41	0.00	0.49	1.00	0.00
Condition moderate	Binary	0.02	0.00	0.13	1.00	0.00
Condition bad	Binary	0.00	0.00	0.04	1.00	0.00
Basement	Binary	0.19	0.00	0.39	1.00	0.00
No basement	Binary	0.81	1.00	0.39	1.00	0.00
Owner-occupied & Non-owner-occupied	Binary	0.17	0.00	0.37	1.00	0.00
Owner-occupied	Binary	0.74	1.00	0.44	1.00	0.00
Non-owner-occupied	Binary	0.09	0.00	0.28	1.00	0.00
Detached house	Binary	0.41	0.00	0.49	1.00	0.00
Non-detached house	Binary	0.59	1.00	0.49	1.00	0.00

3.4 Methodology

3.4.1 Extreme Gradient Boosting – XGBoost

The XGBoost algorithm is chosen as our underlying ML model, since it yielded reasonable results in several research articles (see, e.g., Truong et al., 2020). Especially Stang et al. (2022) showed that the XGBoost achieved the best results for estimating real estate market values in Germany. Therefore, the XGBoost ensures a good model-fit and enables a post-hoc analysis of the results and the application of the PFI and the ALE plots. The XGBoost is a treebased ensemble learning algorithm. Tree-based ensemble learning algorithms combine the results of several decision trees to produce better predictive performance than utilizing a single decision tree. The basic idea behind ensemble learning algorithms is that individual so-called weak learners (e.g., single decision trees) can be combined with each other and thus a strong learner can be achieved. These algorithms were developed to overcome the bias and variance problems associated with single decision trees. In the case of the XGBoost, the following technical expression can be stated:

$$h(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{M} u_m h_m(\boldsymbol{y}|\boldsymbol{x}). \tag{14}$$

where h is the strong learner, h_m are the individual weak learners and u_m is used to weight the weak learners. M is the number of single trees, $\boldsymbol{x}$ is the full features space and $\boldsymbol{y}$ the response variable. Boosting is a type of ensemble learning in which the weak learners h_m are trained sequentially. Starting with one tree, the following models learn from the previous errors. Gradient boosting uses the so-called gradient decent algorithm by adding new trees to minimize the loss of the model. The eXtreme Gradient Boosting is a computationally effective and highly efficient version of gradient boosting. The advantage of XGBoost is that it can recognize very complex patterns within large amount of data. For more information about tree-based methods, ensemble learning and gradient boosting, the interested reader is recommended to Hastie et al. (2001).

3.4.2 Testing Concept

In order to evaluate the XGBoost, five-fold cross validation is used.[16] To obtain the overall performance, we use the set of evaluation metrics presented in Table 3.5. The selected metrics are applied continuously, to evaluate the results of hedonic and machine learning approaches (see, e.g., Mayer et al., 2019).

Table 3.5: Evaluation metrics

Error	Formula	Description
Mean Absolute Percentage Error (MAPE)	$MAPE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert$	Mean of all absolute percentage errors. A lower MAPE signals higher overall prediction accuracy in percent.
Median Absolute Percentage Error (MdAPE)	$MdAPE(y,\hat{y}) = median(\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert)$	Median of all absolute percentage errors. A lower MdAPE denotes a higher precision in percent without being sensitive to outliers.
Error buckets (PE(x))	$PE(x) = 100\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert < x$	Percentage of predictions where the relative deviation is less than x%, with x being 10 and 20. A larger PE(x) signals a lower variation in the predictions.
R^2	$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$	

[16] K-fold cross validation is a method to test how good the predictive power of a statistical model is. It randomly splits the data set into k equal-sized folds (= blocks). One fold is used to test the model, the remaining folds are used for training. This process is performed k times, so that each fold is used once as test data. At the end, the cross-validation error is calculated by averaging the errors of the individual test folds.

	Coefficient of determination. A high R^2 is an indication of better goodness of fit of the model.

3.4.3 Permutation Feature Importance

Permutation feature importance (PFI) is a so-called post-hoc global model-agnostic technique. This term describes the fact that this technique is applied after the actual training of an algorithm (= post-hoc), the results are determined based on all available training observations (= global) and can be applied to different algorithms (= model-agnostic). Starting with the results of a trained valuation algorithm, the PFI can be used to detect the influence of individual features on the target variable of a model. The basic idea behind this technique is that if the values of a given feature are permutated and this particular feature is important for predicting the target variable, then the loss function should increase. The greater the increase, the more important the feature. The result of the PFI analysis is a ranking that indicates the relative importance of all individual features with regard to the ML model. The PFI analysis provides a simple and intuitive visual representation of which factors are important for the algorithm and which play only a subordinate role. We use the PFI to first identify the most important features before analyzing them in more detail afterwards. One main advantage of the PFI is that it can be applied to all machine learning models. Initially introduced only for random forest by Breiman (2001), Fisher et al. (2019) developed this method further to be applicable to all models and called it 'model reliance'. As explained, a feature is considered as important for the final prediction of an ML model, if the prediction error increases after its entries are permuted. Therefore, the permutation feature importance of feature j can be defined as:

$$PFI_j = E\left(L(\hat{f}(\boldsymbol{x}_j, \boldsymbol{x}_{-j}), \boldsymbol{y})\right) - E(L\left(\hat{f}(\boldsymbol{x}, \boldsymbol{y})\right), \tag{15}$$

where L denotes a chosen loss function, $\hat{f}$ refers to a fitted supervised machine learning model, $\boldsymbol{x}_j$ and $\boldsymbol{x}_{-j}$ are the permuted variable j and its complementary set of features. Furthermore, $\boldsymbol{x}$ defines the full features space and $\boldsymbol{y}$ the response variable. In this paper, the Mean Absolute Percentage Error is used as a loss function. For every feature j, the permutation feature importance is computed 100 times, each time randomly permuting its entries. To obtain the final PFI of j, hundred permutation feature importances are averaged.

3.4.4 Accumulated Local Effects

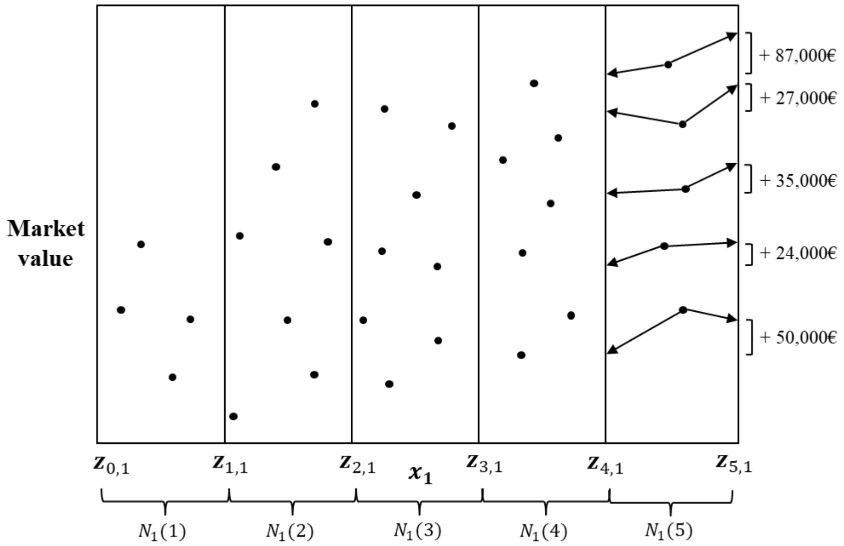
To analyze how a single feature influences the prediction of an ML model on average, so-called global model-agnostic feature effect methods can be used. These methods allow a precise visual representation of the patterns learned by the ML model and tell us how a feature interacts with a target variable. There are different XAI approaches that have been developed to perform such an analysis. To the best of our knowledge, the field of real estate research focused on using Partial Dependence Plots (PDPs) (see, e.g., Lorenz et al., 2022). PDPs are calculated by varying each feature over all observed values (marginal distribution), while holding all other features constant and re-predicting the target variable. The basic idea behind this is that, by varying the inputs and then calculating the outputs, it is possible to analyze how the influence of a feature develops along its actual distribution. By using the marginal distribution, a function is created that is only dependent on the feature of interest. By plotting the average prediction as a function of the respective feature values, the relationships learned by the ML model can be identified and visualized. A detailed description of PDPs can be found in Friedman (2001).

However, the way PDPs are calculated suffers from disadvantages. First, PDPs require a lot of computing time. Second, PDPs assume the relationship between features to be independent, otherwise the PDPs are severely biased. Yet, in the case of real estate, many features are inherently correlated. For example, the living area and the number of rooms are intrinsically interdependent. In the calculation of the PDPs, one would incorporate unrealistic data pairs, such as a house with 400 square meters and 1 room and or a house with 40 square meters and 10 rooms (Molnar, 2020). Therefore, the results of the PDPs should be interpreted with caution.

To overcome this disadvantages, Apley & Zhu (2020) developed Accumulated Local Effects (ALE) plots which attempt to answer the same question as PDPs, namely how features interact with the target variable. But unlike PDPs, ALE plots compute differences in predictions by varying the features value of interest only with closely related data instances (conditional distribution) instead of using the marginal distribution. Accordingly, this time small bins are created for the feature of interest and variations of the feature are made only by means of the upper and lower bounds of these bins. A simple visual representation regarding the logic of ALE plots can be found in Figure 3.1. The figure represents a simple one-dimensional case in which feature x_1 serves as our feature of interest and a total of five bins $N_1(1), \dots, N_1(5)$ are used to separate the dataset.

Figure 3.1: Logic of ALE plots (Adapted from Galkin et al., 2020)

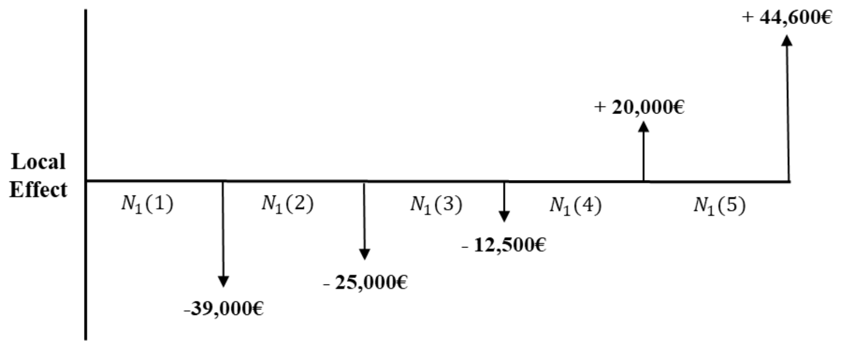
Step 1:



Step 2:

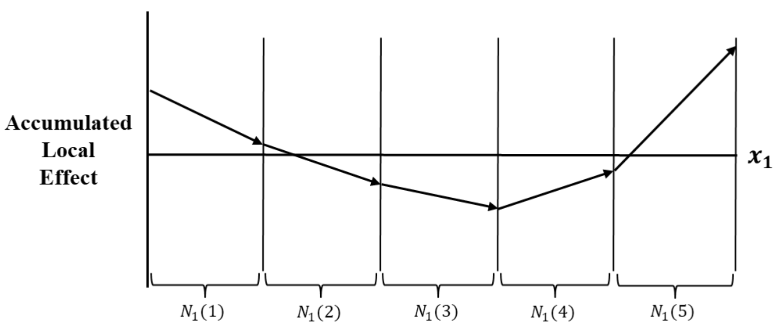
$$\boldsymbol{LE_{N_1(5)}} = \frac{\boldsymbol{(87,000€ + 27,000€ + 35,000€ + 24,000€ + 50,000€)}}{\boldsymbol{5}} = \boldsymbol{+\,44,600€}$$

Step 3:



Step 4:

$$\boldsymbol{ALE_{N_1(1)} = LE_{N_1(1)}}$$
$$\boldsymbol{ALE_{N_1(2)} = LE_{N_1(1)} + LE_{N_1(2)}}$$
$$\dots$$
$$\boldsymbol{ALE_{N_1(5)} = LE_{N_1(1)} + LE_{N_1(2)} + LE_{N_1(3)}, + LE_{N_1(4)}, + LE_{N_1(5)}}$$



The first step of computing the ALE plots is to calculate the so-called Local Effect (LE) of each bin. The LE tells us what effect the feature of interest has on the target variable within a selected bin. For example, in order to build the LE of feature x_1 within the bin $N_1(5)$, we first use the upper bound $z_{5,1}$ for all observations of x_1 within the bin $N_1(5)$ to calculate the prediction of the already trained ML model (while holding the values of all other features constant) and then repeat the same process using the lower bound $z_{4,1}$. In a next step, for all observations of x_1 within the bin $N_1(5)$, the predictions obtained with the lower bound $z_{4,1}$ are subtracted from those with the upper bound $z_{5,1}$ to calculate the change in prediction within the range $N_1(5)$. In the example shown in Figure 3.1, these range from +24,000 € to +87,000 €. Once these steps have been completed for all observations within the bin $N_1(5)$, we calculate the average of all changes to obtain the LE of x_1 within the bin $N_1(5)$, in our example €44,600.

For the final construction of the ALE plots, the individual LEs are accumulated and centered on the basis of the average prediction of all observations. Regarding the accumulation, the LE of the first bin is taken first and then expanded by that of the subsequent bin. This process is repeated until the last bin, and results in a line plot showing the ALEs. The centering of the individual ALEs ensures that the final ALE plot can always be interpreted in such a way that the line shown represents the effect of a feature as a function of the average prediction of the ML model over the full feature space. Therefore, the ALE plots have to be interpreted slightly differently to the PDPs. While PDPs show the average prediction depending on the feature values, ALEs tell us how changes in a feature influence the prediction on average compared to the average prediction. For interested readers, a more technical explanation of how ALE plots work can be found in Appendix II.

In summary, the ALE plots provide us with the opportunity to analyze the features identified in advance by the PFI analysis in more detail and to examine how the inherited relationship is learned by the ML model. By displaying the results via plots, they can be easily and intuitively interpreted. In contrast to the PDPs that have been predominantly used so far, the results of the ALE plots can be assumed to show more accurate and representative results due to the sole use of realistic data pairs. For further information about the ALE plots, we recommend reading Apley & Zhu (2020).

3.5 Results

An extra XGBoost model was trained for each of the seven cities. Furthermore, different algorithms were trained for the condominiums and the single-family homes. To ensure that the machine learning models provide reliable results and can therefore be used as underlying models for the post-hoc XAI analysis, the models were tested by five-fold cross validation on five different evaluation metrics, namely the MAPE, MdAPE, PE(10), PE(20) and R^2. Table 3.6 shows the average results of the evaluation metrics across all cities. To make a statement about the quality of the results, we estimated a basic OLS regression for each city and can thus benchmark the results.[17]

Table 3.6: Results XGBoost for all Top-7 cities

	XGBoost		OLS	
Metrics	**Condominiums**	**Single-family homes**	**Condominiums**	**Single-family homes**
MAPE	0.1253	0.1441	0.1986	0.1834
MdAPE	0.0829	0.0988	0.1467	0.1314
PE(10)	0.5698	0.5073	0.3654	0.3999
PE(20)	0.8247	0.7861	0.6396	0.6833
R^2	0.8117	0.7133	0.6230	0.5623

Our results indicate strong and robust model performance across all Top-7 cities for the XGBoost. For all metrics, the XGBoost yields a better result than the OLS, which could be expected from the literature. The MAPE of the XGBoost, relative to the OLS, is 37% lower for condominiums and 21% lower for single-family homes. The results for each city can be seen in Table 3.9, Appendix III. It is evident that the XGBoost outperforms the basic OLS regression for each city with respect to each evaluation metric used. These differences can already serve as a first indicator of non-linearities, joint effects and higher order interactions within the data. The trained XGBoost algorithms are therefore well suited to a post-hoc analysis of the results and the application of the PFI and the ALE plots.[18]

3.5.1 Results Permutation Feature Importance (PFI)

In a first step, we use the PFI to determine which variables are important for predicting the market value per square meter. The PFI provides a highly compressed, global insight into the machine learning model's behavior. The PFI is easy to interpret and also takes into account interactions within the individual features, as described by Molnar (2020). The PFI ranks all features used in the model according to their influence on the dependent variable. Therefore, in our case, the higher the ranking of a feature, the greater its influence on the

[17] Overall, the magnitude of the improvement over the OLS benchmark may be inflated as it ignores the well-established literature on functional forms of the variables. However, we follow this path for three reasons. First, we want to point out that non-linearities and interactions can imply large performance differences. Second, this baseline OLS will also be used as a benchmark for the ALE plots in the next chapter to emphasize the non-linearity of the data as much as possible and third, by using the ALE plots we want to show how non-parametric ML can help choose suitable functional forms for parametric and semi-parametric models.
[18] To obtain the results of the permutation feature importance and accumulated local effects the scikit-learn (https://scikit-learn.org/stable/modules/permutation_importance.html) and PyALE (https://pypi.org/project /PyALE/) packages are used.

market value per square meter of a property. In Appendix IV, the detailed results of the feature importance analysis can be found. The focus of the PFI analysis is to identify the most important features, which are then used for further investigation using the ALE plots. Therefore, the five most influenceable features for condominiums and single-family homes for each city can be seen in Tables 3.7 and 3.8.

Table 3.7: Top-5 features per city – Condominiums

	Berlin	Hamburg	Munich	Cologne	Frankfurt	Stuttgart	Dusseldorf
Top-1 Feature	Year of valuation	Year of valuation	Year of valuation	Year of valuation	Year of construction	Year of valuation	Year of construction
Top-2 Feature	Year of construction	Unemploy-ment ratio	Year of construction	Year of construction	Year of valuation	Year of construction	Year of Valuation
Top-3 Feature	Unemploy-ment ratio	Year of construction	Longitude	Unemploy-ment ratio	Longitude	Unemploy-ment ratio	Longitude
Top-4 Feature	Longitude	Longitude	Living area	Longitude	Unemploy-ment ratio	Living area	Latitude
Top-5 Feature	Latitude	Latitude	Unemploy-ment ratio	Latitude	Latitude	Longitude	Living area

Table 3.8: Top-5 features per city – Single-family homes

	Berlin	Hamburg	Munich	Cologne	Frankfurt	Stuttgart	Dusseldorf
Top-1 Feature	Year of valuation	Year of valuation	Year of valuation	Year of valuation	Year of valuation	Year of valuation	Lot size
Top-2 Feature	Year of construction	Lot size	Lot size	Living area	Lot size	Lot size	Year of valuation
Top-3 Feature	Lot size	Unemploy-ment ratio	Living area	Lot size	Living area	Living area	Year of construction
Top-4 Feature	Living area	Living area	Year of construction	Unemploy-ment ratio	Longitude	Year of construction	Living area
Top-5 Feature	Longitude	Year of construction	Latitude	Year of construction	Year of construction	Unemploy-ment ratio	Unemploy-ment ratio

In summary, the results of the PFI analysis show the expected results and are also in line with the findings of previous studies, e.g., Cajias (2018) and Lorenz et al. (2022). Thus, it turns out that even with modern ML models, similar to more traditional parametric and semi-parametric approaches, the usual features seem to play a predominant role. The results show that the same features are important for both condominiums and single-family homes. The valuation year is by far the most important. For condominiums, it is always at the top of the list except in Dusseldorf and Frankfurt, and for single-family homes it is also always the highest ranked feature except in Dusseldorf. It shows that the market values of the properties are mainly influenced by the market phase and thus the general market trends. In addition to the valuation year, it can be seen that structural, location-related and socio-economic features have an important influence on the market value of the properties. These findings are also in line with Dubin (1988) and Sirmans et al. (2005). Not only structural characteristics, but also the location and the economic or social environment of a property are decisive for the composition of market values in hedonic pricing models. In the case of condominiums, the year of construction and the living area are the most important factors in terms of property characteristics. In the case of single-family homes, lot size is added to these features. This is also in line with the general findings of other studies (see, e.g., Fan et al., 2006). In our case, the location of the properties is represented by the latitude and longitude. It turns out that depending on the location, the values for otherwise identical properties are different. Socio-economically, we find that the unemployment ratio seems to play a partially important role.

3.5.2 Results Accumulated Local Effects Plots (ALE)

To analyze the identified features in more detail, we use ALE plots to take a closer look at how the individual effects work and what economic insights can therefore be drawn. ALE plots describe the main effect of a feature at a certain point in comparison to the average predicted value. Compared to traditional hedonic price functions, they can capture non-linearities independently, without the need for a priori manual specification. This enables us to visualize a more realistic representation of the actual market fundamentals. Our findings are therefore beneficial for all real estate market players, to reach more targeted and, in particular, data-supported decisions. In addition, the results of the ALE plots can also be used to specify the functional form for each feature for other parametric and semi-parametric models and thus improve the performance of these models as well. Therefore, our results are also important for applications where machine learning is not the method of choice. Researchers and practitioners can use the approximated relations and implement them into their parametric or semi-parametric models.[19] Accordingly, the following results are interesting for a broad audience in real estate research and practice. The ALE plot is centered and the mean effect of the features is zero. Therefore, the y-axis of the ALE plot can be interpreted as the main effect of the independent variable at a certain point, in comparison to the average predicted value. It shows wheter a feature changes, how much it affects the prediction on average compared to the average prediction. The ALE algorithm divides the feature space into intervals containing the same number of data points, whereby feature intervals with a greater observation density are chosen to be smaller than intervals with a low density. In our case, the maximum number

[19] An example of how an OLS can be optimized by using the results of the ALEs can be seen in Appendix V.

of intervals is set to 250. In order to check whether there are non-linearities within the data, we show within the ALE plots, in addition to the effect identified by the XGBoost, the results of a basic OLS as a benchmark. The results of the OLS are shown by means of a red line and show the main difference between previous hedonic pricing approaches and to the results of our analysis. We are thus able to show which effects are covered well by parametric models and which effects must be manually specified.

Starting with the structural characteristics, Figure 3.2 shows the ALE plots for the year of construction for condominiums, and Figure 3.3 for single-family homes. In contrast to the red OLS line, it is obvious that the trend is not linear. The effects are approximately the same across all cities. Comparing the graphs of condominiums with those of single-family homes, we see that the effects are essentially the same. It is notable that the negative trend for middle-aged properties is more pronounced for the former, suggesting that the year of construction has a generally greater influence on condominiums. Traditionally, the effect of the year of construction is described and incorporated as u-shaped (see, e.g., Mayer et al., 2019), as values for new buildings and for old buildings are generally higher than those for middle-aged properties. This effect can be described, for example, by the increased renovation rate for old properties and the generally higher quality of new buildings. While the u-shape transformation seems reasonable in a parametric context, the effects shown in Figures 3.2 and 3.3 indicate that this transformation cannot be supported here. Hence, this transformation can only be seen as a rough approximation of the true underlying relationship. While we also see the effect that middle aged properties tend to have a lower market value than the average valuation, the increase is much more significant for properties with newer construction years than for older buildings. In particular, for properties built between 2010 and 2020, we see that the increase is already almost exponential. This trend reflects the current high demand for new buildings in major German cities. Due to the lack of supply, people are currently willing to pay significantly higher prices for properties on the market. Another explanation lies in the sharp rise of construction costs. The construction of a property has become significantly more expensive in Germany’s metropolitan regions in recent years, which is also ultimately reflected in market values.

Figure 3.2: Condominiums – Year of construction

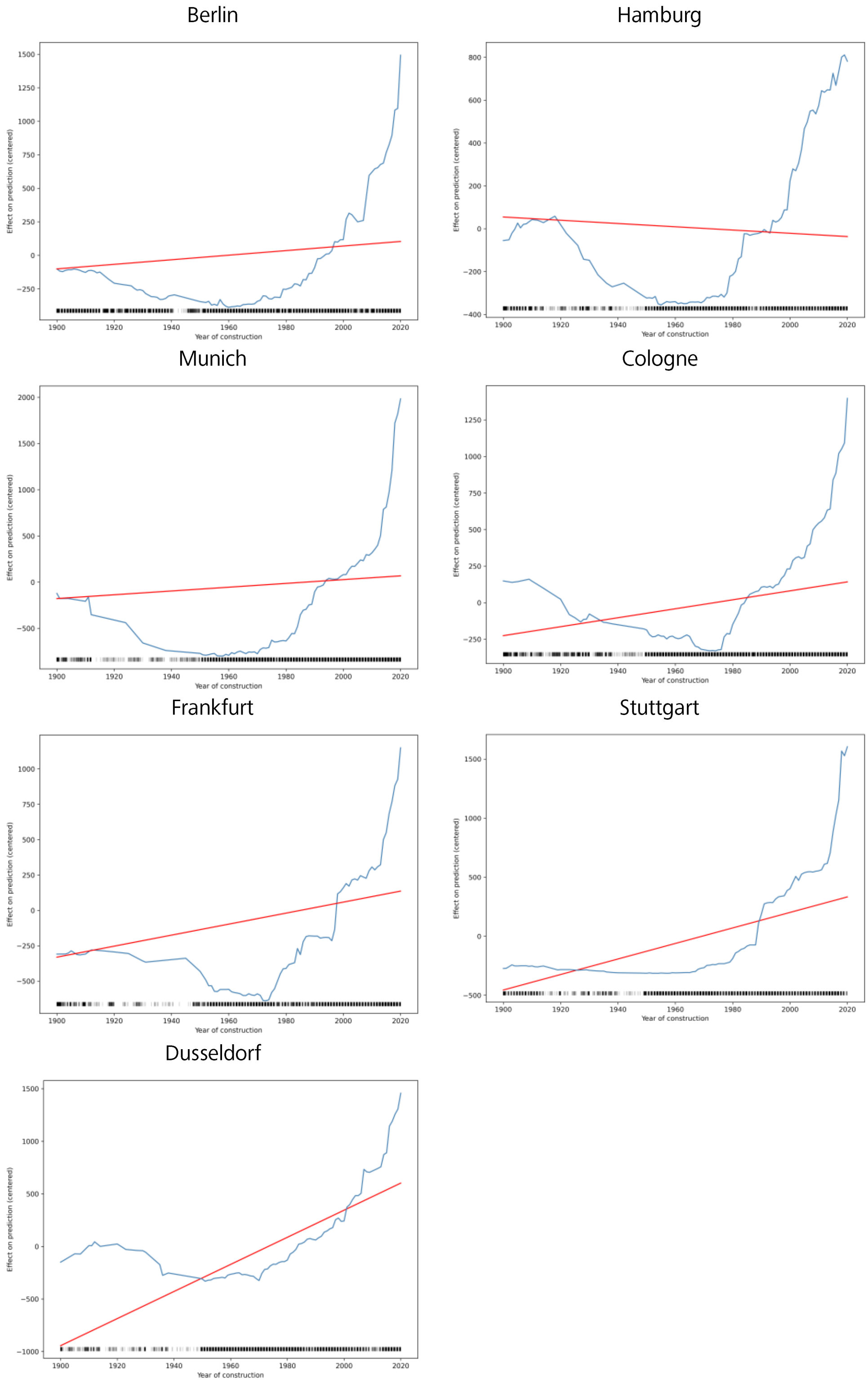
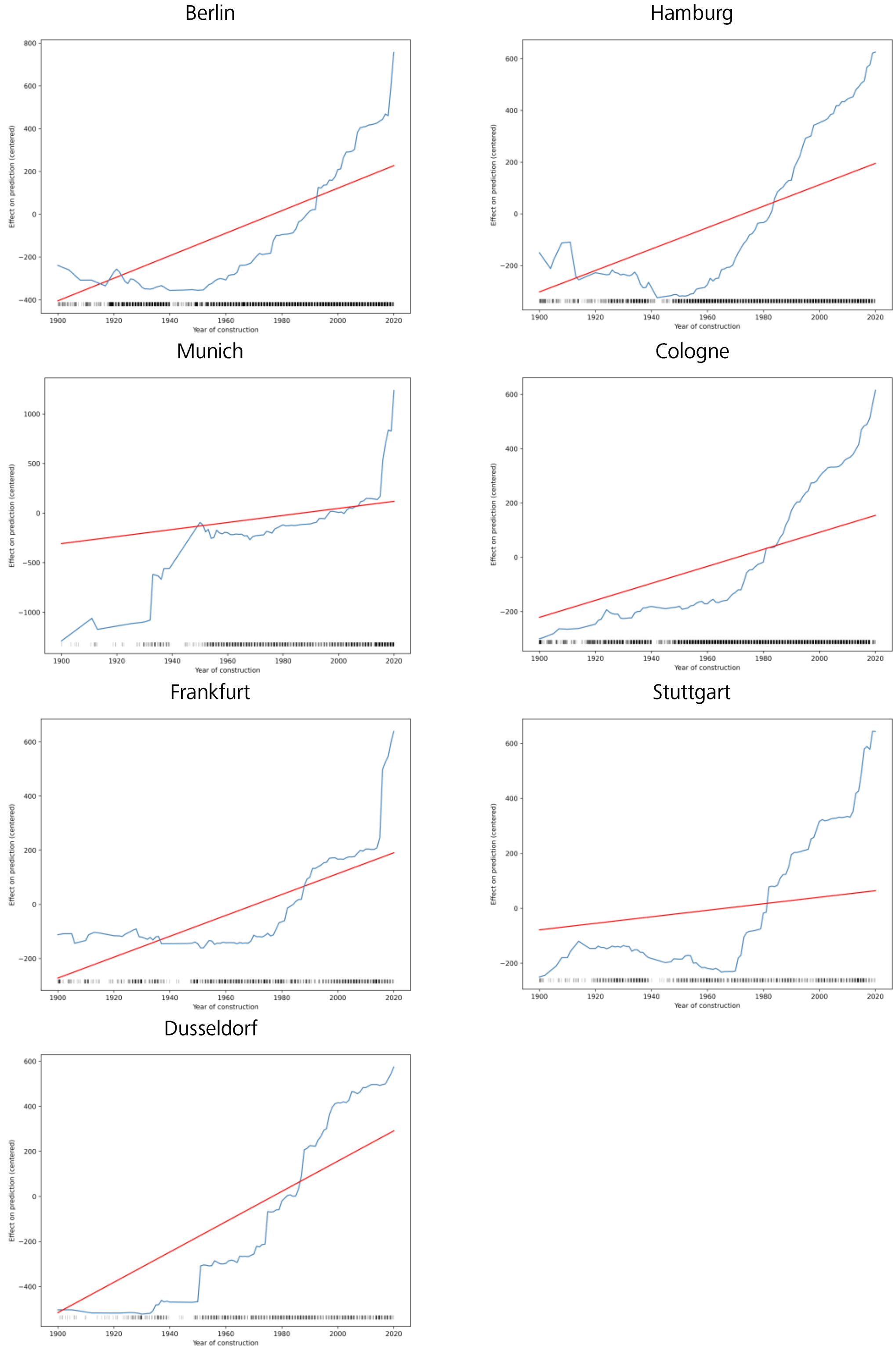


Figure 3.3: Single-family homes – Year of construction



The effect of the living area on market values of condominiums and single-family homes can be seen in Figures 3.4 and 3.5. Focusing on the condominiums, for all Top-7 cities, a clear non-linear relationship is identified with the XGBoost, which would be very difficult to represent by parametric or semi-parametric models. As no recurring pattern is evident, the effects seem to differ in each city. The findings show that there is no generally applicable rule for the analyzed cities and that a well-known rule of thumb in the real estate industry "the larger the area, the lower the market value per square meter" (see, e.g., the results of Wittowsky et al. (2020)) does not hold for condominiums. The ALE plots clearly indicate that there are different patterns regarding the market values within the cities. For example, a high demand for small apartments in the cities of Berlin, Frankfurt, Hamburg, Munich, and Stuttgart is evident. In the cities of Dusseldorf and Cologne, on the other hand, this is not as pronounced. These results offer important implications for the real estate industry. The ALE plots support the analysis of which type of apartment sizes are in demand in which region and what prices can be achieved. Currently, such decisions are often still made on the basis of personal experience or purely descriptive market statistics. The combination of machine learning and ALE plots, on the other hand, enables an empirically valid and data-driven analysis. With respect to the derivation of a generally valid functional form for parametric and semi-parametric models, the results show that this is only possible to a limited extent and that, therefore, in our case an individual function should ideally be selected in each case at the regional level. In contrast to condominiums, the effect of living area on the market value per square meter of single-family homes is homogeneous across all seven cities. Furthermore, the effect is almost linear and can be mapped by the basic OLS to a large extent. Overall, the effect shows a negative trend, which can most likely be attributed to the marginal cost effect. Major components of the costs of a single-family house (e.g., land area, development costs, etc.) are fixed to a certain extent and increase only noticeably as the size of the living area increases. In the case of larger houses, these costs are distributed over the additional square meters and lead to the negative trend shown in Figure 3.5. What also stands out is that the effect for single-family homes is larger on average, which indicates that the size of the property, compared to condominiums, is more important for houses.

Figure 3.4: Condominiums – Living area

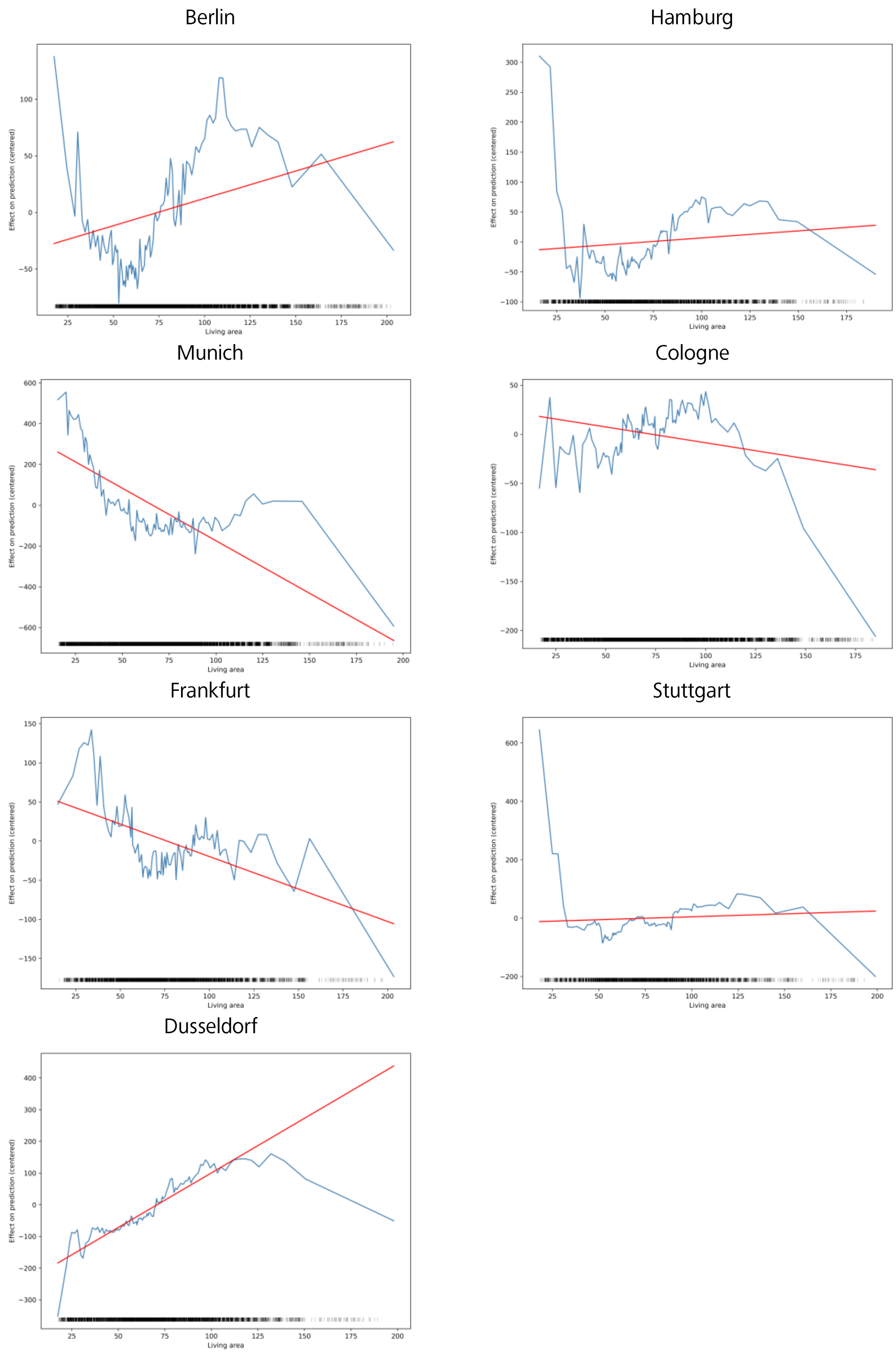


Figure 3.5: Single-family homes – Living area

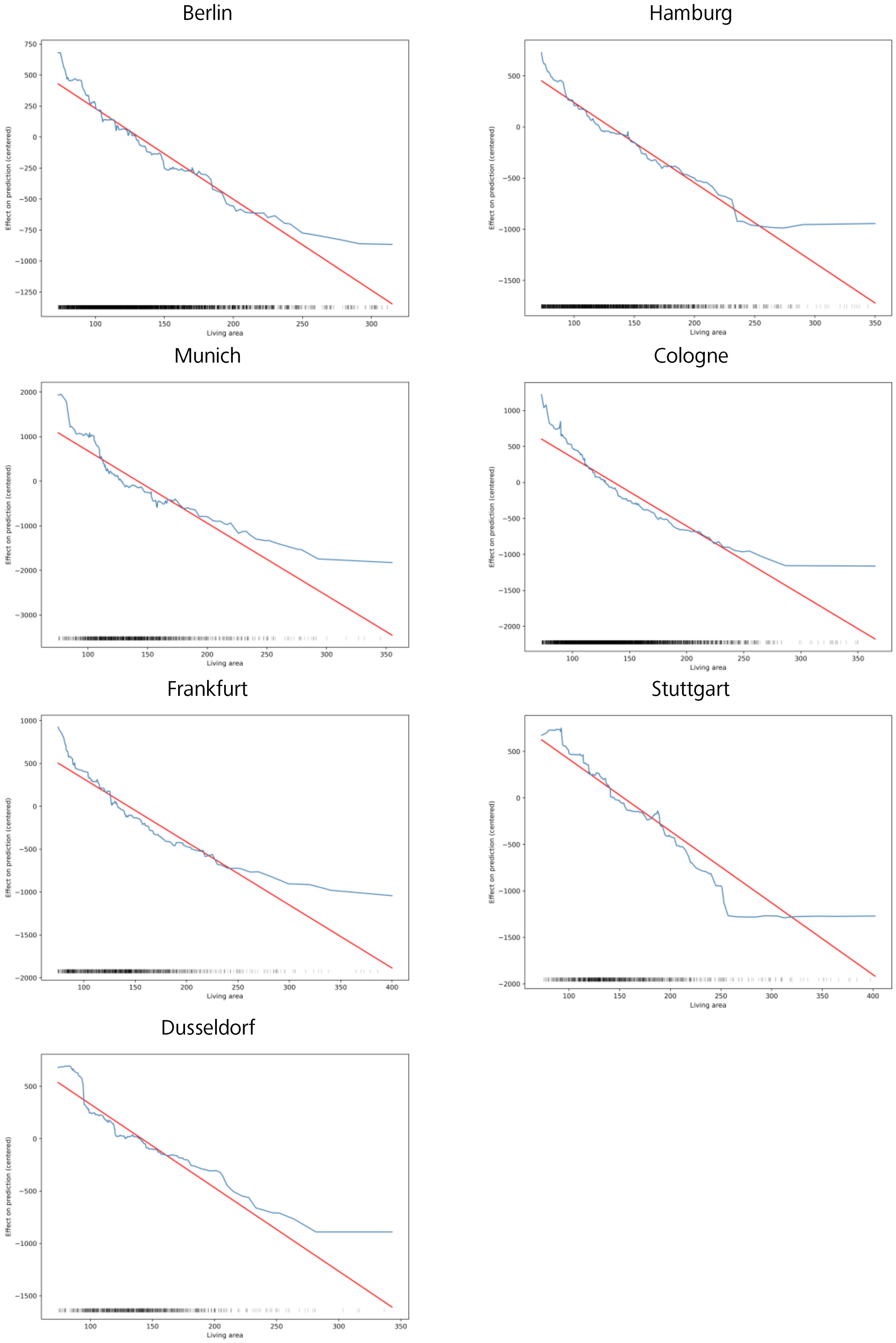


Figure 3.6 highlights the effect of lot size on the market value of single-family homes. In general, the effect is reasonably consistent across all Top-7 cities. The larger the lot size, the higher the market value per square meter of living area. This generally indicates that there is a higher demand for larger plots of land and that market values are rising due to a lack of supply, although a decreasing marginal utility can be seen for very large plots in all cities. While the basic OLS also assumes such a progression, the results of the XGBoost are more granular and thus more accurately reflect the actual effect. In general, it can be seen that a log transformation of the lot size appears to be useful with respect to a manual specification in parametric and semi-parametric approaches. This can also be adopted as generally valid for all cities analyzed in our case. Differences between cities can be seen mainly in the strength of the effect. In Munich and Stuttgart, for example, the lot size seems to play a more important role than in Berlin or Hamburg.

As the results of the PFI show, the year of valuation was by far the most important feature. The effect of the valuation year is shown in Figure 3.7 for condominiums, and in Figure 3.8 for single-family houses. Since this variable is discrete, the next lower and next higher values are used as interval limits. The bars represent the size of the sample in each year, and the number is summarized with a second y-axis on the right of the plot. The red line again represents the results of a basic OLS. The results show that in all Top-7 cities, market values have risen sharply and constantly over the observation period. In principle, the OLS and XGBoost curves are relatively similar. However, it is apparent that the XGBoost identifies a stronger price increase for the last three years. In general, the demand for both condominiums and single-family homes has risen sharply in German metropolitan areas. Since supply is inelastic because of long development periods, this increase in demand leads to a dynamic rise in prices (Belke & Keil, 2018). Our findings show that this price increase also affects the market values of the properties. Property values have risen over the past few years, irrespective of their structural and locational characteristics. This decoupling effect can be seen as quite critical, as the generally strongly rising prices can lead to speculation, which was also observable on the U.S. residential real estate markets before the Global Financial Crisis (Martin, 2011). In combination with a longer time series and other important macroeconomic features, the ALE plots could be used to conduct a more in-depth analysis and thus analyze key developments and drivers of real estate price bubbles. We consider this to be a promising and interesting area of research that should be pursued further.

Figure 3.6: Single-family homes – Lot size

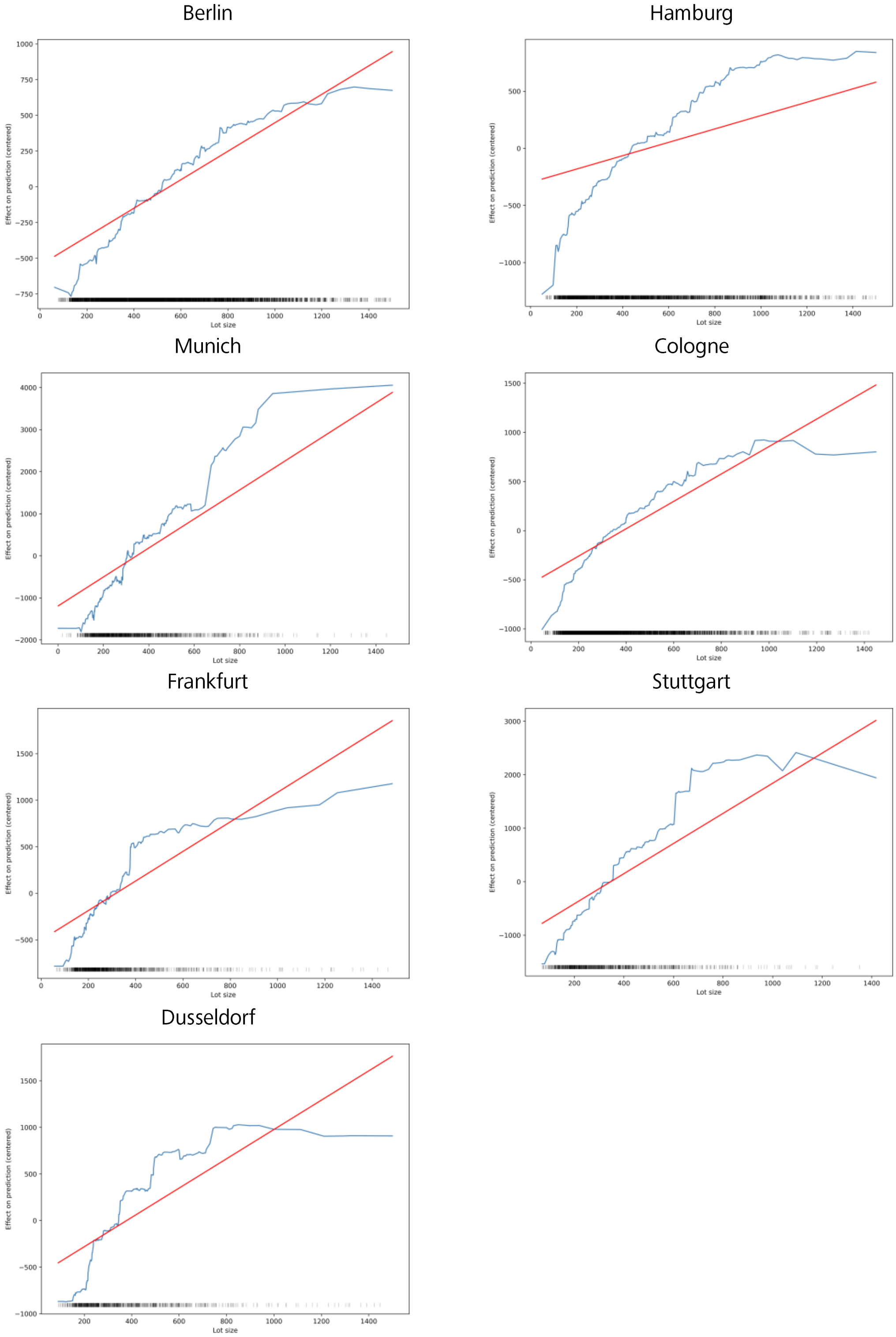


Figure 3.7: Condominiums – Year of valuation

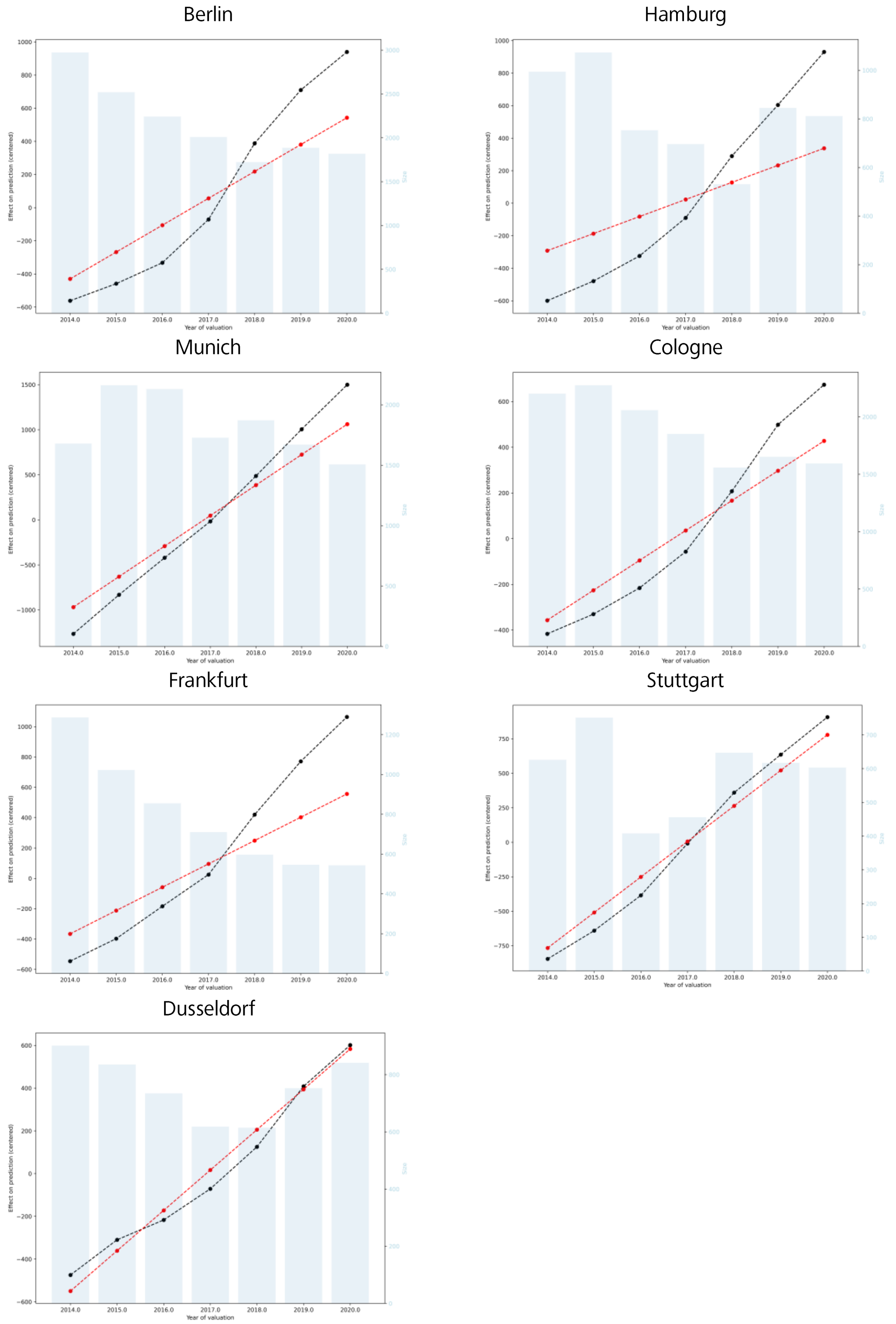
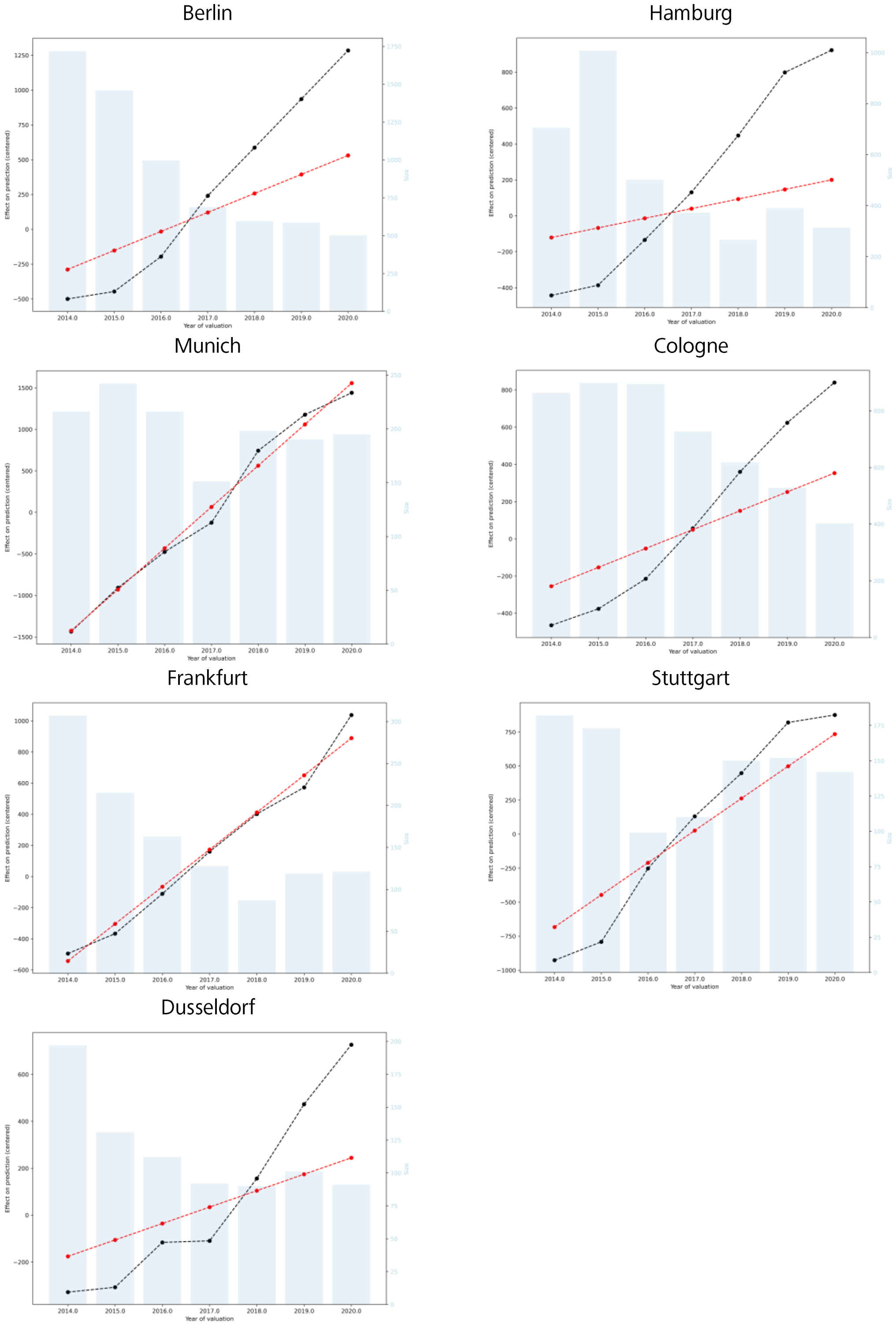


Figure 3.8: Single-family homes – Year of valuation



In addition to the temporal and structural characteristics, the PFI analysis shows that the location-related features of latitude and longitude play a significant role in predicting the market values per square meter for both condominiums and single-family homes. The effect of the latitude for condominiums is shown in Figure 3.9 and the effect of the longitude in Figure 3.10. In both figures, one can see that prices are rising sharply in certain regions – the city centers. An exception to this is the latitude of the city of Stuttgart. This is due to the unique location of Stuttgart, in a valley with a lot of industry in central locations and thus has its own geographical characteristics. Comparing the ALE plots of the XGBoost and the basic OLS, it can be clearly stated that a simple OLS cannot reproduce these non-linear locational effects. It is also not possible to derive a universally valid functional form, which is why in our case a specification at the local level is also recommended. Looking at the impact of latitude and longitude on single-family houses in Figures 3.11 and 3.12, a clear difference can be seen. Not only is the effect much less pronounced, but the expensive regions are no longer in the center of the cities, which is not surprising, since the houses in these cities are located in the suburbs. In summary, the ALE plots of latitude and longitude can help to identify promising locations within the cities.[20]

Finally, we consider the impact of the unemployment ratio on market values. Figure 3.13 shows the impact on condominiums and Figure 3.14 on single-family homes. Overall, the ALE plots of the XGBoost seem reasonable and are in line with the findings of other studies (see, e.g., Grum & Govekar, 2016). In all cities, the XGBoost identifies a downward trend in market values the higher the unemployment rate. Comparing the results of the XGBoost with the baseline OLS, one can see a large difference between the two graphs. In this context, the focus is on Stuttgart in particular, where positive effect of the OLS is assumed. Once again, the OLS is not able to capture the effects in a granular and comprehensive way. Decisions made on the basis of this flawed assumption can have far-reaching consequences and should be avoided.[21]

[20] However, these results have to be interpreted with caution. There are several location-based features in our dataset. Besides latitude and longitude, there are the four micro-scores, which also describe the surrounding location of the properties. Furthermore, there are three socioeconomic variables in the dataset, which are available at the ZIP code level and thus could also be seen as a proxy for location. To obtain the overall effect of the location on the price, these individual effects would have to be aggregated.

[21] However, it bears repeating at this point that the results should also be interpreted with caution, as the unemployment rate can also serve as a simple proxy for the location of a property, due to its availability at the ZIP code level.

Figure 3.9: Condominiums – Latitude

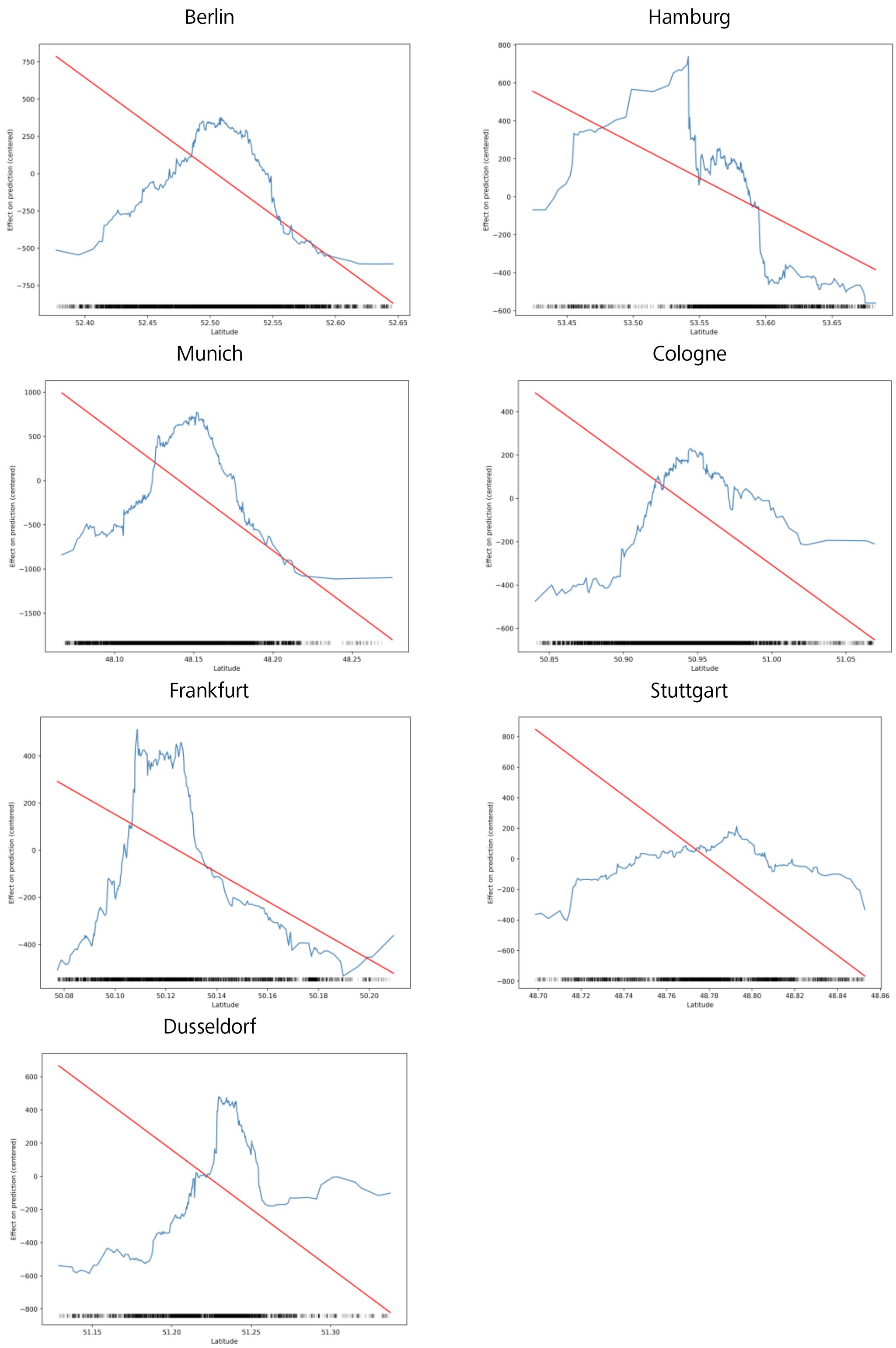


Figure 3.10: Condominiums – Longitude

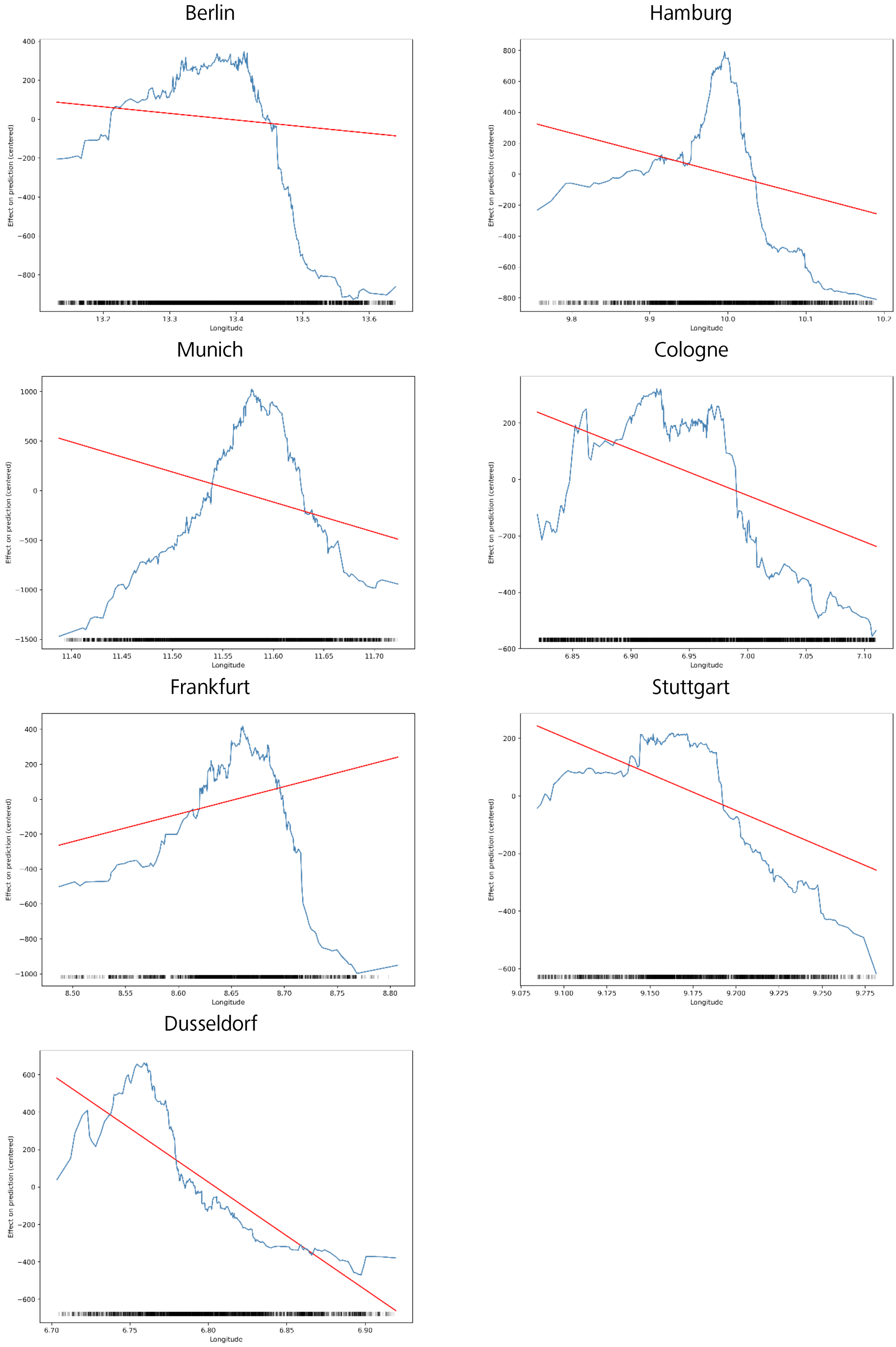


Figure 3.11: Single-family homes – Latitude

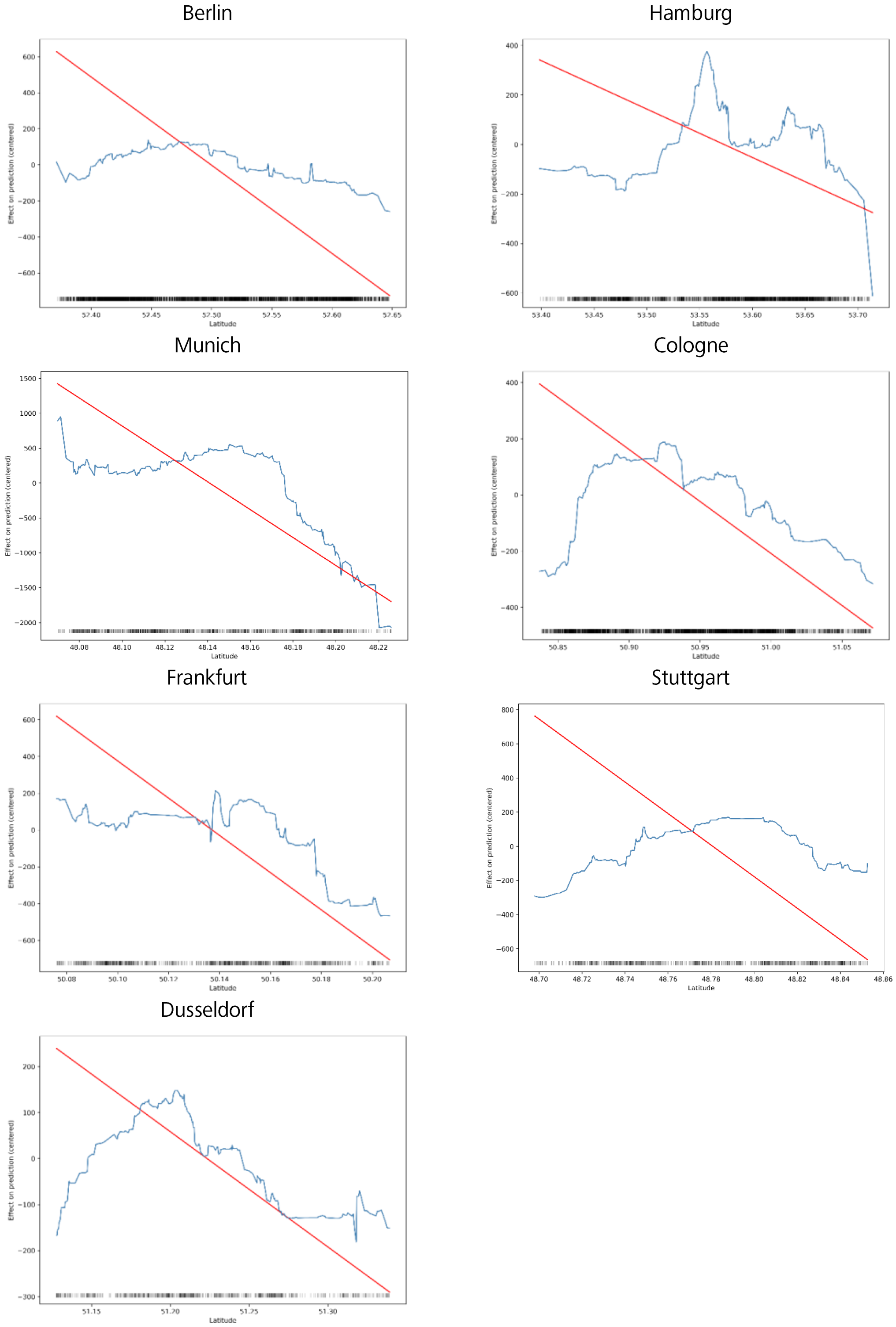


Figure 3.12: Single-family homes – Longitude

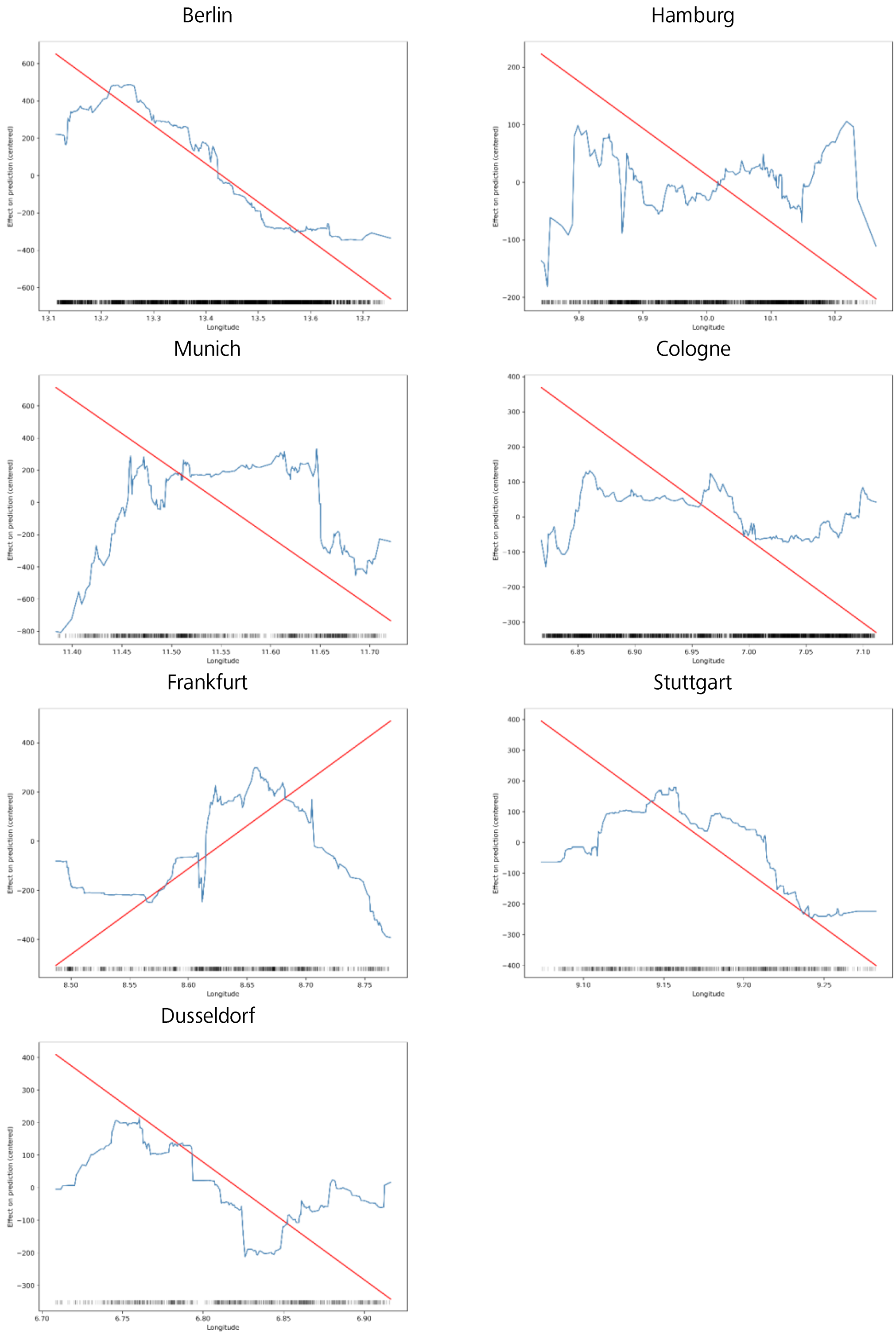


Figure 3.13: Condominiums – Unemployment ratio

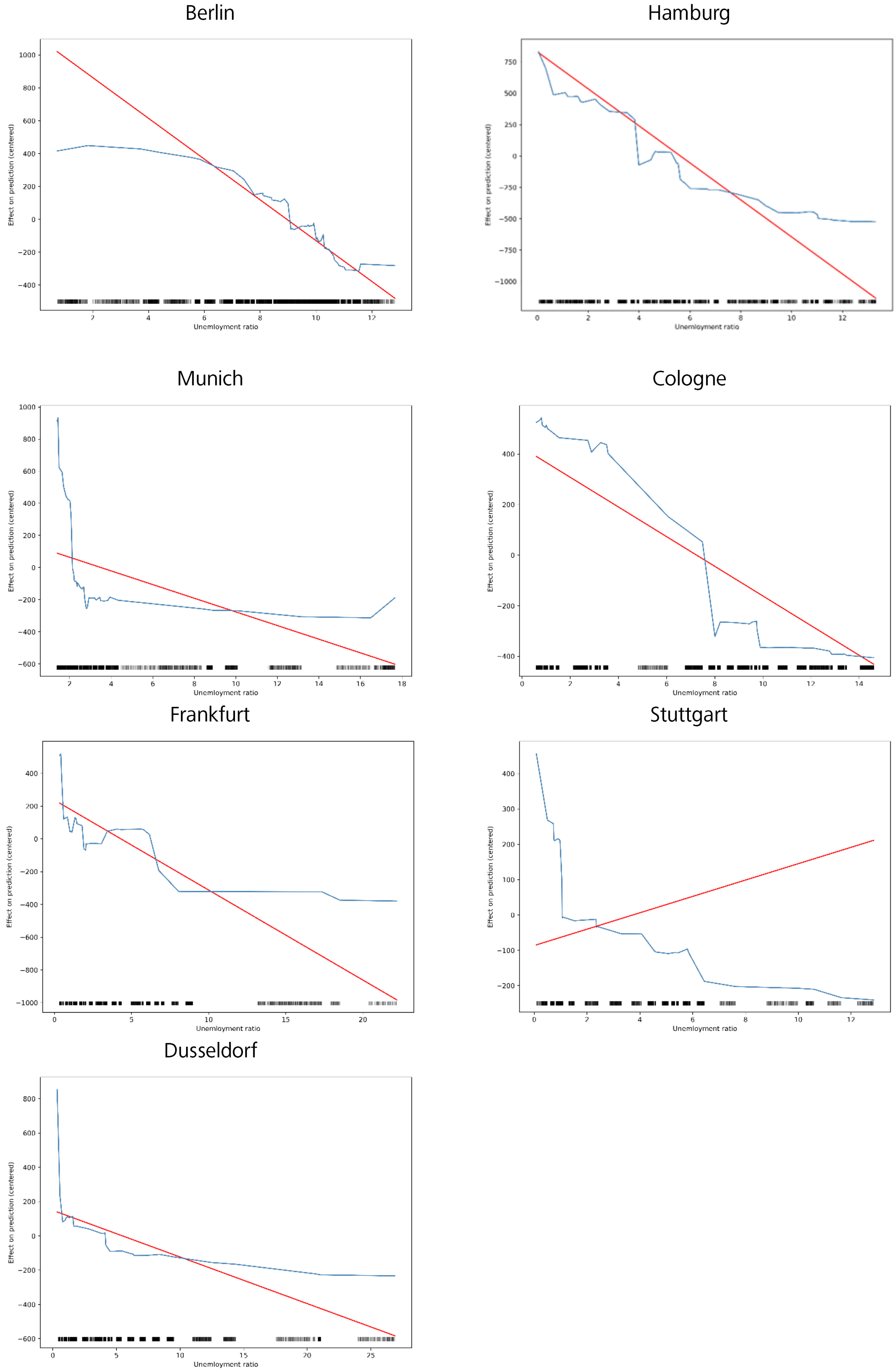
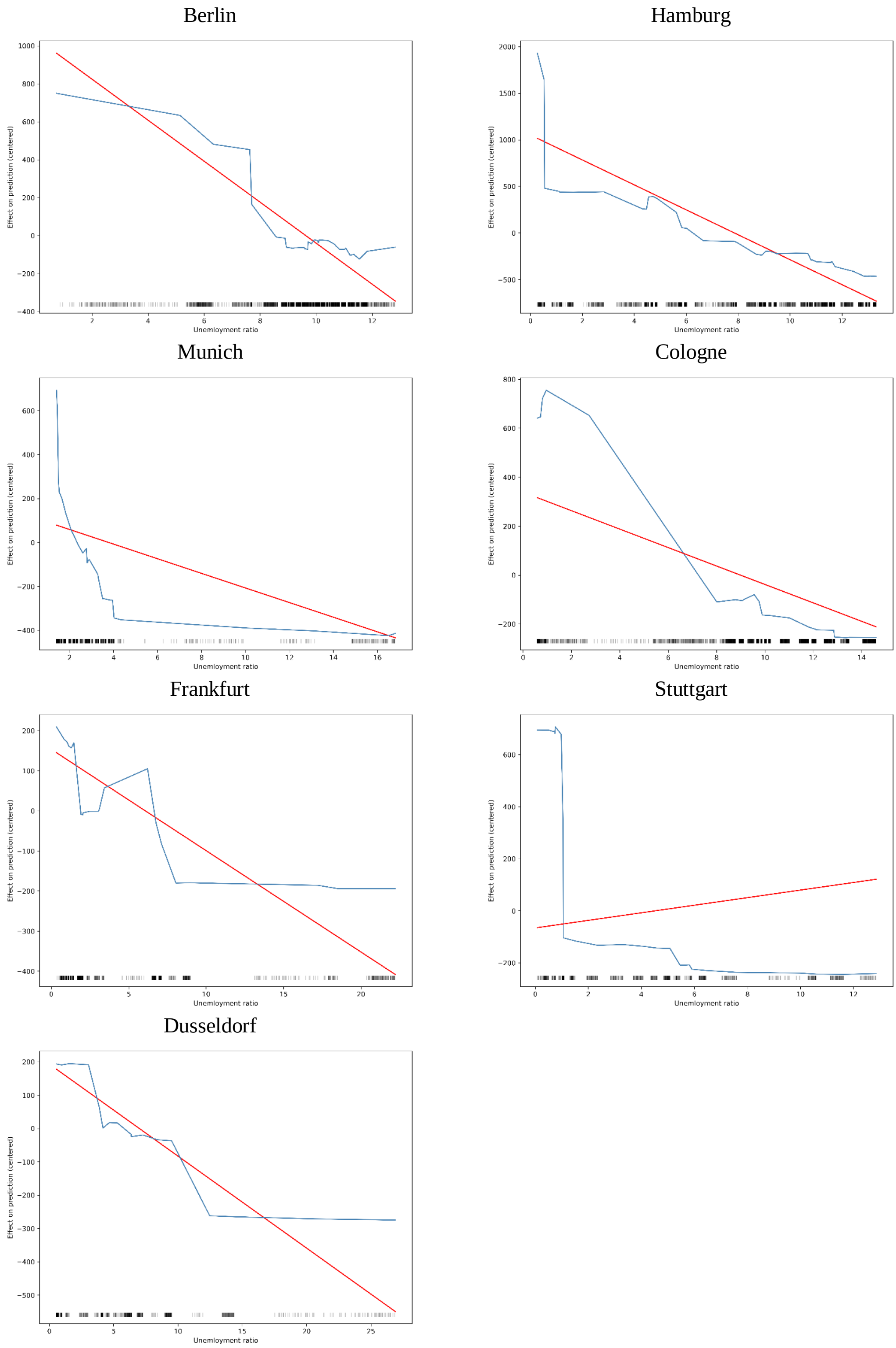


Figure 3.14: Single-family homes – Unemployment ratio



3.6 Conclusion

This study is intended to introduce eXplainable Artificial Intelligence (XAI) in a real estate context, and updates the existing literature with the application of Accumulated Local Effects plots (ALE). Compared to the Partial Dependence Plots (PDP), which are commonly used in real estate research to date, ALE plots can also handle correlated features and are therefore more suitable for real-estate-related problems. We use a dataset consisting of 61,763 condominiums and 19,403 single-family homes for the Top-7 cities of Germany to empirically apply our analysis. We are interested in identifying the most value-determining features of the two property subtypes, and then analyzing them separately with respect to existing non-linearities. We apply Permutation Feature Importance (PFI) to identify the most important features, and ALE plots to visualize their individual effects. As an underlying machine learning (ML) model, we use the XGBoost algorithm for a hedonic estimation of the properties' market values.

The PFI analysis shows that the same features play a predominant role for both condominiums and single-family homes. The valuation year is by far the most important feature. In addition to the valuation year, a mix of structural, location-related and socio-economic features influence the market value of the properties. Among the structural characteristics, the most important are the year of construction, the living area and the lot size. In terms of location features, the latitude and longitude are decisive in terms of market values. Socio-economically, we find that the unemployment ratio seems to play a partially important role.

To the best of our knowledge, we are the first to use ALE plots for visualizing individual effects on the market value, and we see that both non-linear and linear effects can be observed. In terms of year of construction, our results show that for both condominiums and single-family homes, the u-shaped transformation traditionally used for Hedonic Price Models (HPMs) is not evident. Properties with newer construction years are valued much higher than is the case for older buildings. We can confirm that properties with middle age tend to have lower market values. The results for living area among condominiums are particularly interesting. The ALE plots show no clear trend here for the cities studied, but that this effect varies greatly and is clearly non-linear. For single-family homes, on the other hand, a linear trend can be observed for all cities. An approximately linear trend is also evident for the year of valuation feature. The analysis of the ALE plots of latitude and longitude shows that market values within the city can vary greatly, depending on the particular location. Our results show that market values for condominiums tend to be highest in the centers of cities, whereas values for single-family homes tend to be highest outside city centers. Both findings appear intuitive and are in line with the prevailing opinion within the real estate industry. The effect of the unemployment rate is also clearly non-linear and different across the cities analyzed. In general, however, the presumed negative influence is evident.

In summary, the ALE plots provide a deeper understanding of the fundamentals of real estate markets and either empirically confirm long-established rules of thumb or, as in the case of living area for condominiums, challenge them. Our results show that linear relationships indeed occur in the housing market. Here, parametric estimates can also provide valuable results. However, the analysis of the features year of construction, living area, lot size, latitude, longitude and unemployment ratio reveal non-linear effects. Therefore, non-parametric ML approaches seem to be the right choice. The ALE plots offer a way to represent these effects in a well-founded way and thus make an important contribution to the housing market literature. Moreover, the effects seen in the ALE plots can be used to optimize parametric and semi-parametric models in order to achieve a higher predictive performance.

Model-agnostic methods are still a rather young field of research, but will play a major role in the acceptance of ML methods in the future, as they allow us to look into the "black box "of ML approaches and are thus an important tool in deciphering them. The tradeoff between explainability and model performance can thus be mitigated in the long run. However, further research is still needed before widespread use is possible. With regard to future work, for example, a more in-depth analysis of neighborhood and environmental features constitutes a promising field of application. While the focus of this study was rather on the analysis of structural characteristics, the analysis of neighborhood and environmental features in the context of property prices has become an important part of real estate research and can, in our opinion, also benefit from the advantageous properties of XAI approaches and especially those of ALE plots.

3.7 Appendix

3.7.1 Appendix I – Micro Score

Our gravity model can be described using an activity function $f(A_p)$ and a distance function $f(D_{i,p})$:

$$A_{i,p} = \sum f(A_p) f(D_{i,p}). \tag{16}$$

$A_{i,p} \in [0,100]$ denotes the accessibility of point i for the POI p, whereby the activity function $f(A_p)$ specifies the relative importance of POI p, with $f(A_p) \in [0,1]$. $f(D_{i,p})$ measuring the travel time from point i to the POI p by using a non-symmetric sigmoidal distance function. The travel time was obtained for the selected POIs via Open Street Map, and normalized using the following function:

$$L(x) = \frac{K}{(1 + Qe^{0.5x})^{\frac{1}{v}}}, \tag{17}$$

where $K, Q \in \mathbb{R}$ and $v \in \mathbb{R}^{+}$ are defined for all possible distances $x \in \mathbb{R}$. Furthermore, we have:

$$\begin{gathered} K = (1 + Q)^{1+v}, \\ Q = v \cdot \exp(B \cdot x^{*}), \\ v = \frac{\exp(B \cdot x^{*}) - 1}{\ln(y_i) - 1}, \end{gathered} \tag{18}$$

where x^{*} denotes a feature specific point of inflection and y^{*} is 0.5.

3.7.2 Appendix II – Accumulated Local Effects

Accumulated local effects (ALE), developed by Apley & Zhu (2020), is a feature effect approach that shows how a feature influences a prediction on average. The technical explanation of how this can be achieved is given in the following section.

Assume that $\boldsymbol{x}$ is the full feature space containing d variables and $\boldsymbol{y}$ the response variable. $\hat{f}$ is a fitted supervised machine learning model that is differentiable and uses $\boldsymbol{x}$ to predict $\boldsymbol{y}$. Define $\boldsymbol{x}_j$ as the feature of interest and $\boldsymbol{x}_{-j}$ the complement set of features, $j \in \{1, \dots d\}$. X_j represents the j^{th} feature as a random variable. Then, the ALE main effect of $\boldsymbol{x}_j$ can be calculated:

$$\hat{f}_{j,ALE} = \int_{z_{0,j}}^{x_j} E\left[\frac{\partial \hat{f}(X_j, X_{-j})}{\partial X_j} \mid X_j = z_j\right] dz_j - constant, \tag{19}$$

with $z_{0,j}$ being a lower bound of X_j. Usually, $z_{0,j}$ is defined as $\min\{\boldsymbol{x}_j\}$. The expected value E is computed conditional on the representation of $\boldsymbol{x}_j$ and over the marginal distribution of $\boldsymbol{x}_{-j}$. The constant is subtracted to center the plot.

Since not every machine learning model is differentiable, Apley & Zhu (2020) introduced a way to estimate the ALE for any supervised machine learning model. Therefore, the value range of the j^{th} feature is divided into K intervals $N_j(k)$, $\{N_j(k) = (z_{k-1,j}, z_{k,j}]: k = 1,2, \dots K\}$, where $z_{j,k}$ refers to the upper and $z_{k-1,j}$ the lower boundary of interval k. Furthermore, x^* is a specific value of $\boldsymbol{x}_j$ and $k_j(x^*)$ denotes the index of the interval x^* belongs to. $n_j(k)$ is the number of observations in each interval k and $\boldsymbol{x}_{i,-j}$ represents the observations of the remaining features, $i \in \{1, 2, \dots N\}$.

Before we can compute the main effect $\hat{f}_{j,ALE}$, the uncentered ALE $\hat{g}_{j,ALE}$ of the j^{th} feature has to be calculated for every $x^* \in (z_{0,j}, z_{K,j}]$, where $z_{0,j}$ is just below the minimum observation of $\{x_{i,j}: i = 1, \dots n\}$ and $z_{K,j}$ is the maximum observation of $\{x_{i,j}: i = 1, \dots n\}$:

$$\hat{g}_{j,ALE}(x^*) = \sum_{k=1}^{k_j(x^*)} \frac{1}{n_j(k)} \sum_{\{i: x_{i,j} \in N_j(k)\}} \left[\hat{f}(z_{k,j}, \boldsymbol{x}_{i,-j}) - \hat{f}(z_{k-1,j}, \boldsymbol{x}_{i,-j})\right]. \tag{20}$$

Now the ALE main effect estimator can be computed by subtracting an estimate of $E[g_{j,ALE}(X_j)]$:

$$\begin{aligned} \hat{f}_{j,ALE}(x^*) &= \hat{g}_{j,ALE}(x^*) - \frac{1}{n}\sum_{i=1}^{n} \hat{g}_{j,ALE}(x_{i,j}) \\ &= \hat{g}_{j,ALE}(x^*) - \frac{1}{n}\sum_{k=1}^{K} n_j(k) \cdot \hat{g}_{j,ALE}(z_{k,j}). \end{aligned} \tag{21}$$

3.7.3 Appendix III – Evaluation Metrics at City Level

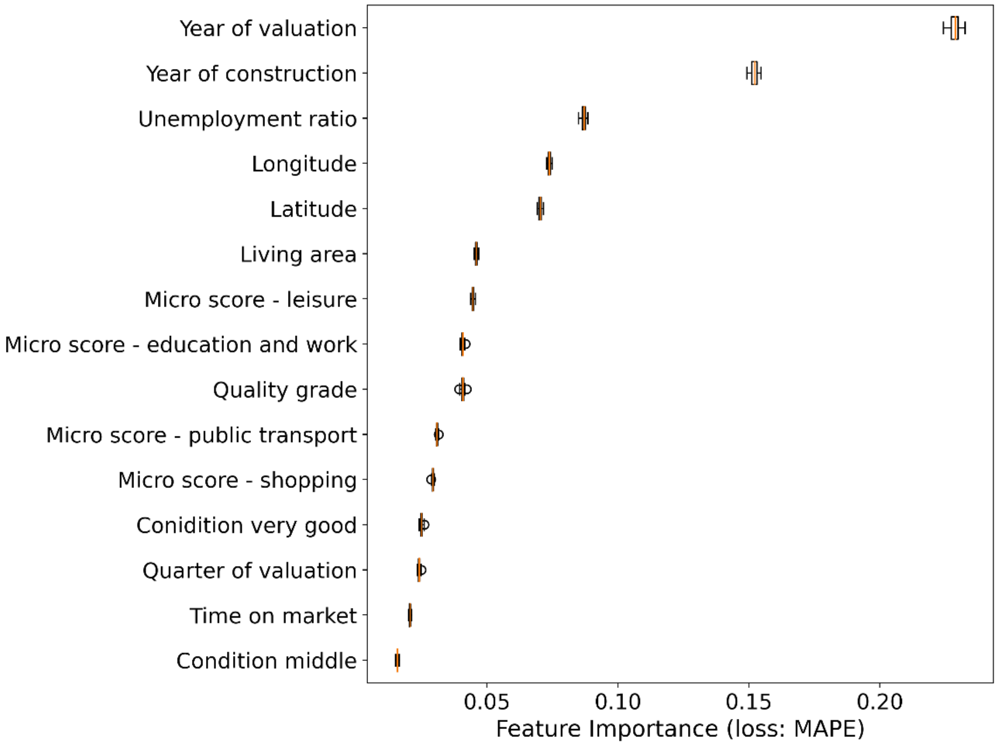
Table 3.9: XGBoost results at city level

	XGBoost		OLS	
Metrics	**Condominiums**	**Single-family homes**	**Condominiums**	**Single-family homes**
		Berlin		
MAPE	0.1431	0.1415	0.2311	0.1837
MdAPE	0.0984	0.0988	0.1741	0.1342
PE(10)	0.5059	0.5062	0.3077	0.3927
PE(20)	0.7752	0.7910	0.5593	0.6698
R^2	0.8052	0.7544	0.6192	0.6074
		Hamburg		
MAPE	0.1291	0.1505	0.1990	0.2039
MdAPE	0.0806	0.1047	0.1455	0.1466
PE(10)	0.5697	0.4805	0.3721	0.3595
PE(20)	0.8199	0.7710	0.6369	0.6450
R^2	0.7936	0.7245	0.6123	0.5288
		Munich		
MAPE	0.1051	0.1735	0.1718	0.2016
MdAPE	0.0618	0.0981	0.1233	0.1251
PE(10)	0.6559	0.5099	0.4163	0.4154
PE(20)	0.8772	0.7670	0.7104	0.7095
R^2	0.8079	0.6381	0.5734	0.5264
		Cologne		
MAPE	0.1278	0.1232	0.2008	0.1587
MdAPE	0.0865	0.0878	0.1469	0.1180
PE(10)	0.5577	0.5530	0.3609	0.4381
PE(20)	0.8267	0.8220	0.6398	0.7217
R^2	0.8388	0.7256	0.6820	0.5672
		Frankfurt		
MAPE	0.1121	0.1571	0.2124	0.1866
MdAPE	0.0738	0.1041	0.1571	0.1222
PE(10)	0.6080	0.4816	0.3360	0.4351
PE(20)	0.8512	0.7640	0.6050	0.7070
R^2	0.8312	0.6639	0.6061	0.5431
		Stuttgart		
MAPE	0.1160	0.1513	0.1449	0.1859
MdAPE	0.0824	0.1033	0.1082	0.1337
PE(10)	0.5772	0.4851	0.4698	0.3720
PE(20)	0.8508	0.7778	0.7607	0.6845
R^2	0.8002	0.6869	0.6855	0.5518
		Dusseldorf		
MAPE	0.1334	0.1848	0.1907	0.2019
MdAPE	0.0930	0.1270	0.1438	0.1450
PE(10)	0.5296	0.4158	0.3622	0.3624
PE(20)	0.7921	0.6695	0.6440	0.6474
R^2	0.7793	0.4909	0.5867	0.4188

3.7.4 Appendix IV – Permutation Feature Importance

To get a better understanding on how to interpret the PFI results, Figure 3.15 shows the PFI plot including the 14 most important features for condominiums in Munich.

Figure 3.15: Condominiums – PFI feature importance – Berlin



The features are ranked on the y-axis from most important to least important. The x-axis provides information of how much the average prediction accuracy changes when the values of the features are permutated 100 times. The individual feature importance ratios in Figures 3.15 can be interpreted like box plots. The orange line in the middle represents the median values and the ends of the bars the 25% and 75% quantiles. Furthermore the 1.5x interquartile range (IQR) and individual outliers are shown.

The year of valuation is seen to have the highest impact on valuation accuracy for condominiums in Berlin. By randomly permuting the year of valuation 100 times, the models MAPE increases by 23.0% for condominiums, while the other features are kept constant. The second most important feature with an increase of 15.3% in the MAPE is the year of construction for condominiums, followed by the unemployment ratio.

The remaining PFI plots in Figures 3.16 and 3.17 can be interpreted in the same way.

Figure 3.16: Feature importance – Condominiums

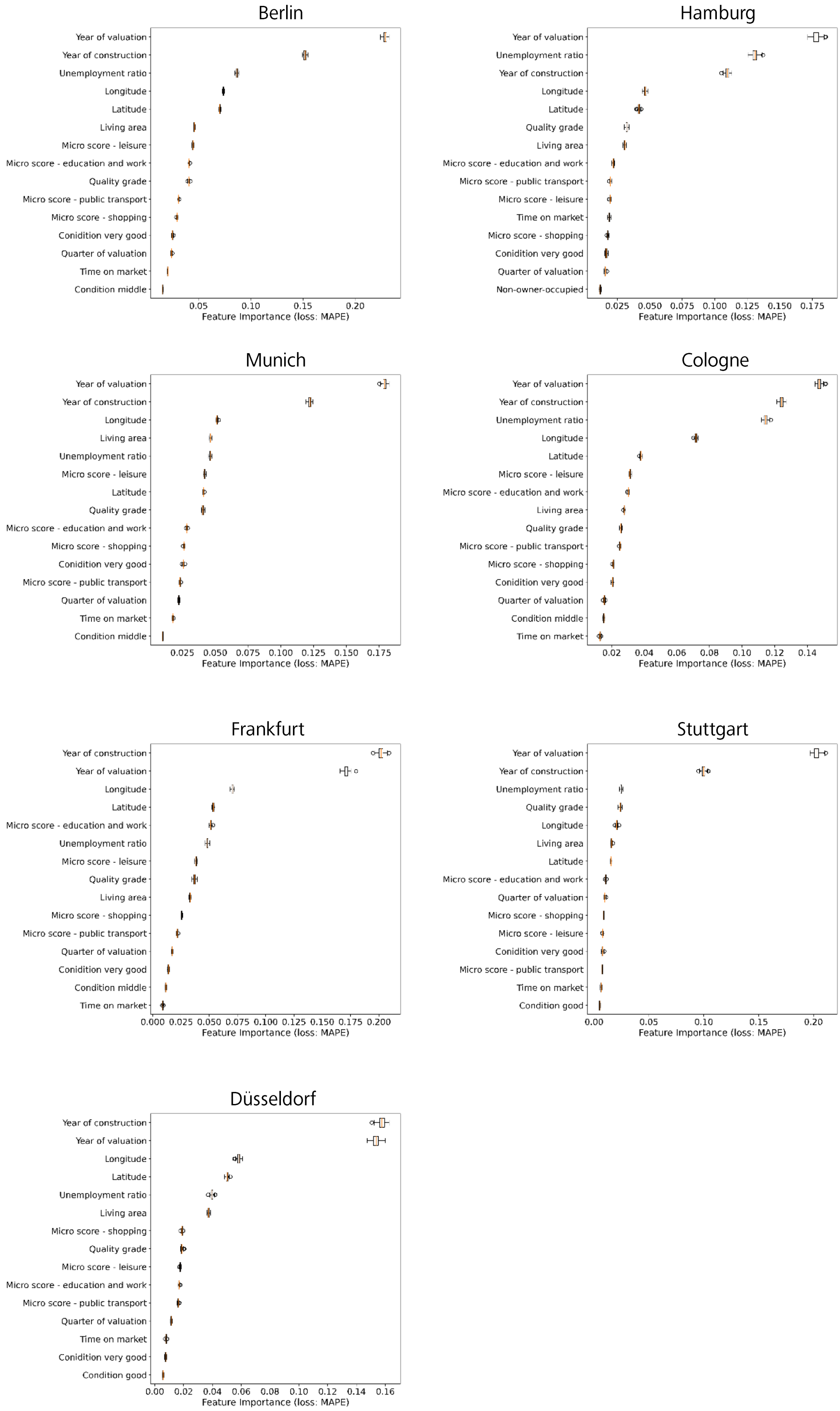
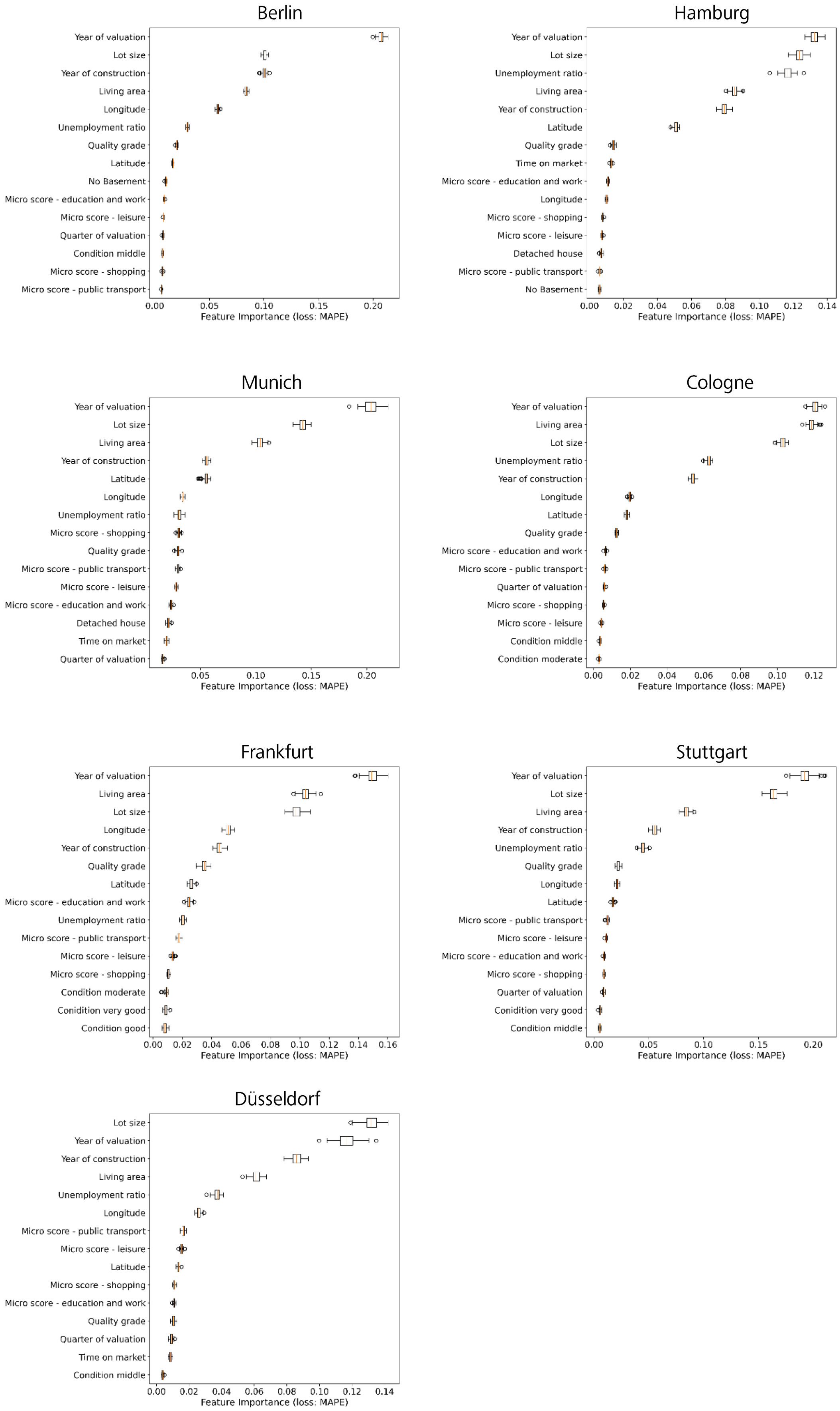


Figure 3.17: Feature importance – Single-family homes



3.7.5 Appendix V – Using ALE Plots to Optimize an Ordinary Least Square Regression

Below, it is shown how, for example, the results of the ALEs can be used to optimize the benchmark OLS used in this paper to achieve higher predictive performance. To do this, as the first step, the ALE plots for each city for condominiums and single-family homes have to be analyzed and relevant inflections of the function have to be found. Then, the feature space of the OLS is split according to those functions in order to better adapt the data to the non-linearities found. Table 3.10 shows the exact feature splits. If we take the year of construction of single-family houses in Berlin as an example, this means that instead of one feature with a feature space of 1900-2020, we now have three features, each with feature spaces of 1900-1953, 1954-2016, and 2016-2020.

Table 3.10: Feature space splits

	Condominiums	Single-family homes
Berlin		
Year of construction	$x \leq 1960$ $1960 < x \leq 2008$ $2008 < x$	$x \leq 1953$ $1953 < x \leq 2016$ $2016 < x$
Year of valuation	$x \leq 2016$ $2016 < x$	$x \leq 2016$ $2016 < x$
Latitude	$x \leq 52.52$ $52.52 < x$	$x \leq 52.59$ $52.59 < x$
Longitude	$x \leq 13.415$ $13.415 < x \leq 13.75$ $13.75 < x$	$x \leq 13.25$ $13.25 < x \leq 13.51$ $13.51 < x$
Unemployment ratio	$x \leq 7.7$ $7.7 < x \leq 11$ $11 < x$	$x \leq 7.68$ $7.68 < x \leq 8.01$ $8.01 < x$
Living area	$x \leq 62.5$ $62.5 < x \leq 110$ $110 < x$	$x \leq 150$ $150 < x \leq 200$ $200 < x$
Lot size	-	$x \leq 500$ $500 < x$
Hamburg		
Year of construction	$x \leq 1918$ $1918 < x \leq 1955$ $1955 < x \leq 1977$ $1977 < x$	$x \leq 1941$ $1941 < x \leq 1968$ $1968 < x$
Year of valuation	$x \leq 2016$ $2016 < x$	$x \leq 2015$ $2015 < x \leq 2018$ $2018 < x$
Latitude	$x \leq 53.545$ $53.545 < x \leq 53.60$ $53.60 < x$	$x \leq 53.515$ $53.515 < x \leq 53.555$ $53.555 < x \leq 53.675$ $53.765 < x$
Longitude	$x \leq 9.975$ $9.975 < x \leq 10$ $10 < x \leq 10.05$ $10.05 < x$	$x \leq 9.8$ $9.8 < x \leq 10.16$ $10.16 < x$
Unemployment ratio	$x \leq 6$ $6 < x$	$x \leq 0.8$ $0.8 < x$
Living area	$x \leq 35$ $35 < x \leq 100$ $100 < x$	$x \leq 150$ $150 < x \leq 235$ $235 < x$
Lot size	-	$x \leq 400$ $400 < x \leq 400$ $800 < x$

	Munich	
Year of construction	$x \leq 1952$ $1952 < x \leq 1980$ $1980 < x \leq 2009$ $2009 < x$	$x \leq 1950$ $1950 < x \leq 2015$ $2015 < x$
Year of valuation	$x \leq 2017$ $2017 < x$	$x \leq 2017$ $2017 < x$
Latitude	$x \leq 48.15$ $48.15 < x \leq 48.21$ $48.21 < x$	$x \leq 48.165$ $48.165 < x$
Longitude	$x \leq 11.58$ $11.58 < x \leq 11.66$ $11.66 < x$	$x \leq 11.465$ $11.465 < x \leq 11.647$ $11.647 < x$
Unemployment ratio	$x \leq 3$ $3 < x$	$x \leq 4$ $4 < x$
Living area	$x \leq 60$ $60 < x \leq 150$ $150 < x$	$x \leq 135$ $135 < x \leq 225$ $225 < x$
Lot size	-	$x \leq 510$ $510 < x \leq 930$ $930 < x$
	Cologne	
Year of construction	$x \leq 1977$ $1977 < x \leq 2007$ $2007 < x$	$x \leq 1970$ $1970 < x \leq 2000$ $2000 < x$
Year of valuation	$x \leq 2017$ $2017 < x$	$x \leq 2016$ $2016 < x$
Latitude	$x \leq 50.895$ $50.895 < x \leq 50.947$ $50.947 < x \leq 51.02$ $51.02 < x$	$x \leq 50.875$ $50.875 < x \leq 50.935$ $50.935 < x$
Longitude	$x \leq 6.93$ $6.93 < x \leq 6.98$ $9.98 < x \leq 7.015$ $7.015 < x$	$x \leq 6.865$ $6.865 < x \leq 6.995$ $6.995 < x$
Unemployment ratio	$x \leq 8$ $8 < x$	$x \leq 8$ $8 < x$
Living area	$x \leq 100$ $100 < x \leq 138.5$ $138.5 < x$	$x \leq 140$ $140 < x \leq 250$ $250 < x$
Lot size	-	$x \leq 300$ $300 < x \leq 950$ $950 < x$
	Frankfurt	
Year of construction	$x \leq 1945$ $1945 < x \leq 1975$ $1975 < x \leq 2009$ $2009 < x$	$x \leq 1980$ $1980 < x \leq 2015$ $2015 < x$
Year of valuation	$x \leq 2017$ $2017 < x$	$x \leq 2017$ $2017 < x$
Latitude	$x \leq 50.112$ $50.112 < x \leq 50.132$ $50.132 < x \leq 50.138$ $50.143 < x$	$x \leq 50.16$ $50.16 < x$
Longitude	$x \leq 8.68$ $8.68 < x \leq 8.74$ $8.74 < x$	$x \leq 8.57$ $8.57 < x \leq 8.66$ $8.66 < x$
Unemployment ratio	$x \leq 9$ $9 < x$	$x \leq 2$ $2 < x \leq 8$ $8 < x$
Living area	$x \leq 55$ $55 < x \leq 150$ $150 < x$	$x \leq 150$ $150 < x \leq 250$ $250 < x$
Lot size	-	$x \leq 400$ $400 < x$

Stuttgart		
Year of construction	$x \leq 1967$ $1967 < x \leq 1988$ $1988 < x \leq 2009$ $2009 < x$	$x \leq 1970$ $1970 < x \leq 2009$ $2009 < x$
Year of valuation	$x \leq 2016$ $2016 < x$	$x \leq 2015$ $2015 < x$
Latitude	$x \leq 48.75$ $48.75 < x \leq 48.75$ $48.79 < x$	$x \leq 48.81$ $48.81 < x$
Longitude	$x \leq 9.15$ $9.15 < x \leq 9.19$ $9.19 < x$	$x \leq 9.165$ $9.165 < x$
Unemployment ratio	$x \leq 1$ $1 < x$	$x \leq 1.5$ $1.5 < x$
Living area	$x \leq 30$ $30 < x \leq 125$ $125 < x$	$x \leq 150$ $250 < x$
Lot size	-	$x \leq 680$ $680 < x$
Dusseldorf		
Year of construction	$x \leq 1971$ $1971 < x \leq 2009$ $2009 < x$	$x \leq 1952$ $1952 < x$
Year of valuation	$x \leq 2016$ $2016 < x$	$x \leq 2017$ $2017 < x$
Latitude	$x \leq 51.20$ $51.20 < x$	$x \leq 51.21$ $51.21 < x$
Longitude	$x \leq 6.76$ $6.76 < x \leq 6.785$ $6.785 < x$	$x \leq 6.765$ $6.765 < x$
Unemployment ratio	$x \leq 1.3$ $1.3 < x$	$x \leq 12.5$ $12.5 < x$
Living area	$x \leq 117$ $117 < x$	$x \leq 175$ $175 < x$
Lot size	-	$x \leq 790$ $790 < x$

To test wheter the splits improved the model performance, we again used five-fold cross validation with the same evaluation metrics as before. Table 3.11 shows the overall results of the optimized OLS with the basic OLS as a benchmark. We can see that the feature splits clearly improved the OLS regarding all evaluation metrics for both condominiums and single-family homes.

Table 3.11: Optimized OLS results for all Top-7 cities

	$OLS_{optimized}$		**OLS**	
Metrics	**Condominiums**	**Single-family homes**	**Condominiums**	**Single-family homes**
MAPE	0.1841	0.1705	0.1986	0.1834
MdAPE	0.1345	0.1243	0.1467	0.1314
PE(10)	0.3934	0.4149	0.3654	0.3999
PE(20)	0.6780	0.7072	0.6396	0.6833
R^2	0.6626	0.6148	0.6230	0.5623

The results on a city level are shown in Table 3.12. The optimized OLS outperforms the basic OLS for all cities regarding all evaluation metrics.

Table 3.12: Optimized OLS results at city level

	$OLS_{optimized}$		OLS	
Metrics	**Condominiums**	**Single-family homes**	**Condominiums**	**Single-family homes**
		Berlin		
MAPE	0.2216	0.1723	0.2311	0.1837
MdAPE	0.1659	0.1268	0.1741	0.1342
PE(10)	0.3233	0.4138	0.3077	0.3927
PE(20)	0.5816	0.6956	0.5593	0.6698
R^2	0.6462	0.6512	0.6192	0.6074
		Hamburg		
MAPE	0.1823	0.1824	0.1990	0.2039
MdAPE	0.1279	0.1301	0.1455	0.1466
PE(10)	0.4037	0.3911	0.3721	0.3595
PE(20)	0.6910	0.6879	0.6369	0.6450
R^2	0.6644	0.6512	0.6123	0.5288
		Munich		
MAPE	0.1633	0.1813	0.1718	0.2016
MdAPE	0.1171	0.1213	0.1233	0.1251
PE(10)	0.4388	0.4173	0.4163	0.4154
PE(20)	0.7350	0.7171	0.7104	0.7095
R^2	0.5921	0.6003	0.5734	0.5264
		Cologne		
MAPE	0.1839	0.1509	0.2008	0.1587
MdAPE	0.1304	0.1140	0.1469	0.1180
PE(10)	0.3924	0.4444	0.3609	0.4381
PE(20)	0.6863	0.7418	0.6398	0.7217
R^2	0.7260	0.6068	0.6820	0.5672
		Frankfurt		
MAPE	0.1810	0.1796	0.2124	0.1822
MdAPE	0.1344	0.1237	0.1571	0.1247
PE(10)	0.3876	0.4210	0.3360	0.4188
PE(20)	0.6761	0.6951	0.6050	0.6886
R^2	0.6835	0.5683	0.6061	0.5601
		Stuttgart		
MAPE	0.1310	0.1748	0.1449	0.1859
MdAPE	0.0971	0.1317	0.1082	0.1337
PE(10)	0.5120	0.3863	0.4698	0.3720
PE(20)	0.8060	0.6949	0.7607	0.6845
R^2	0.7430	0.6038	0.6855	0.5518
		Dusseldorf		
MAPE	0.1734	0.1994	0.1907	0.2019
MdAPE	0.1325	0.1420	0.1438	0.1450
PE(10)	0.3914	0.3476	0.3622	0.3624
PE(20)	0.6862	0.6698	0.6440	0.6474
R^2	0.6344	0.4442	0.5867	0.4188

3.8 References

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.

Belke, A., & Keil, J. (2018). Fundamental Determinants of Real Estate Prices: A Panel Study of German Regions. *International Advances in Economic Research*, *24*(1), 25–45.

Bogin, A. N., & Shui, J. (2020). Appraisal Accuracy and Automated Valuation Models in Rural Areas. *The Journal of Real Estate Finance and Economics*, *60*(1-2), 40–52.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial Dependence, Housing Submarkets, and House Price Prediction. *The Journal of Real Estate Finance and Economics*, *35*(2), 143–160.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2010). Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions. *Journal of Real Estate Research*, *32*(2), 139–159.

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

Brunauer, W. A., Lang, S., Wechselberger, P., & Bienert, S. (2010). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *The Journal of Real Estate Finance and Economics*, *41*(4), 390–411.

Cajias, M. (2018). Is there room for another hedonic model? *Journal of European Real Estate Research*, *11*(2), 204–245.

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.

Cheng, A. C. S., & Fung, M. K. (2015). Determinants of Hong Kong's Housing Prices. *Journal of Economics, Business and Management*, *3*(3), 352–355.

Chiarazzo, V., Caggiani, L., Marinelli, M., & Ottomanelli, M. (2014). A Neural Network based Model for Real Estate Price Estimation Considering Environmental Quality of Property Location. *Transportation Research Procedia*, *3*, 810–817.

Chun Lin, C., & Mohan, S. B. (2011). Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis*, *4*(3), 224–243.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental Variables and Real Estate Prices. *Urban Studies*, *38*(11), 1989–2000.

Dubin, R. A. (1988). Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms. *The Review of Economics and Statistics*, *70*(3), 466.

Fan, G.-Z., Ong, S. E., & Koh, H. C. (2006). Determinants of House Price: A Decision Tree Approach. *Urban Studies*, *43*(12), 2301–2315.

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.

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

Galkin, F., Aliper, A., Putin, E., Kuznetsov, I., Gladyshev, V. N., & Zhavoronkov, A. (2020). Human Gut Microbiome Aging Clock Based on Taxonomic Profiling and Deep Learning. iScience, 23(6), 101199.

Grum, B., & Govekar, D. K. (2016). Influence of Macroeconomic Factors on Prices of Real Estate in Various Cultural Environments: Case of Slovenia, Greece, France, Poland and Norway. *Procedia Economics and Finance*, *39*, 597–604.

Gu, G., & Xu, B. (2017). Housing Market Hedonic Price Study Based on Boosting Regression Tree. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *21*(6), 1040–1047.

Handy, S. L., & Clifton, K. J. (2001). Evaluating Neighborhood Accessibility: Possibilities and Practicalities. *Journal of Transportation and Statistics*, *4*(2-3), 67-78.

Hastie, T., Friedman, J. & Tibshirani, R. (2001). *The Elements of Statistical Learning.* New York: 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.

Huang, Y., & Dall'erba, S. (2021). Does Proximity to School Still Matter Once Access to Your Preferred School Zone Has Already Been Secured? *The Journal of Real Estate Finance and Economics*, *62*(4), 548–577.

Kagie, M., & van Wezel, M. (2007). Hedonic price models and indices based on boosting applied to the Dutch housing market. *Intelligent Systems in Accounting, Finance & Management*, *15*(3-4), 85–106.

Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal of Portfolio Management*, *43*(6), 202–211.

Levantesi, S., & Piscopo, G. (2020). The Importance of Economic Variables on London Real Estate Market: A Random Forest Approach. *Risks*, *8*(4), 112.

Lorenz, F., Willwersch, J., Cajias, M., & Fuerst, F. (2022). Interpretable machine learning for real estate market analysis. Journal of *Real Estate Economics*, Forthcoming.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. *Housing Economics and Public Policy*, 1, 67–89.

Martin, R. (2011). The local geographies of the financial crisis: From the housing bubble to economic recession and beyond. *Journal of Economic Geography*, *11*(4), 587–618.

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.

Metzner, S., & Kindt, A. (2018). Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties. *International Journal of Housing Markets and Analysis*, *11*(1), 73–100.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

Nobis, C. & Kuhnimhof, T. (2018). *Mobilität in Deutschland – MiD: Ergebnisbericht*. Bundesministerium für Verkehr und digitale Infrastruktur, Bonn.

Pace, R. K., & Hayunga, D. (2020). Examining the Information Content of Residuals from Hedonic and Spatial Models Using Trees and Forests. *The Journal of Real Estate Finance and Economics*, *60*(1-2), 170–180.

Peterson, S., & Flanagan, A. (2009). Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research*, *31*(2), 147–164.

Powe, N. A., Garrod, G. D., & Willis, K. G. (1995). Valuation of urban amenities using an hedonic price model. *Journal of Property Research*, *12*(2), 137–147.

Sangani, D., Erickson, K. & Hasan, M.A. (2017). *Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting*. In IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Orlando.

Schnaidt, T., & Sebastian, S. (2012). German valuation: Review of methods and legal framework. *Journal of Property Investment & Finance*, *30*(2), 145–158.

Schulz, R., Wersing, M., & Werwatz, A. (2014). Automated valuation modelling: a specification exercise. *Journal of Property Research*, *31*(2), 131–153.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, *13*(1), 1–44.

Stang, M., Krämer, B., Nagl, C., & Schäfers, W. (2022). From human business to machine learning—Methods for automating real estate appraisals and their practical implications. *Zeitschrift Für Immobilienökonomie*, Advanced online publication.

Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, *174*, 433–442.

van Wezel, M., Kagie, M., & Potharst, R. (2005). Boosting the accuracy of hedonic pricing models. *Report / Econometric Institute, Erasmus University Rotterdam* (EI 2005-50).

Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. (2020). Residential housing prices: Impact of housing characteristics, accessibility and neighbouring apartments – a case study of Dortmund, Germany. *Urban, Planning and Transport Research*, *8*(1), 44–70.

Worzala, E., Lenk, M., & Silva, A. (1995). An Exploration of Neural Networks and Its Application to Real Estate Valuation. *Journal of Real Estate Research*, *10*(2), 185–201.

Yang, J., Bao, Y., Zhang, Y., Li, X., & Ge, Q. (2018). Impact of Accessibility on Housing Prices in Dalian City of China Based on a Geographically Weighted Regression Model. *Chinese Geographical Science*, *28*(3), 505–515.

4 Changing the Location Game – Improving Location Analytics with the Help of Explainable AI

4.1 Abstract

Besides its structural and economic characteristics, the location of a property is probably one of the most important determinants of its underlying value. In contrast to property valuations, there are hardly any approaches to date that evaluate the quality of a real estate location in an automated manner. The reasons are the complexity, the number of interactions and the non-linearities underlying the quality specifications of a certain location. These are difficult to represent by traditional econometric models. The aim of this paper is thus to present a newly developed data-driven approach for the assessments of real estate locations. By combining a state-of-the-art machine learning algorithm and the local post-hoc model agnostic method of Shapley Additive Explanations, the newly developed SHAP location score is able to account for empirical complexities, especially for non-linearities and higher order interactions. The SHAP location score represents an intuitive and flexible approach based on econometric modeling techniques and the basic assumptions of hedonic pricing theory. The approach can be applied post-hoc to any common machine learning method and can be flexibly adapted to the respective needs. This constitutes a significant extension of traditional urban models and offers many advantages for a wide range of real estate players.

Keywords: Location Analytics, Explainable AI, Machine Learning, Shapley Values, Automated Location Valuation Model

4.2 Introduction

Besides its structural and economic characteristics, the location of a property is probably one of the most important determinants of its underlying value. While there have been standardized and globally accepted methods for assessing property values (see, e.g., Parker, 2016), the assessment of the location quality at a given site remains a kind of arbitrary estimation based on more or less subjective individual opinions. Interestingly, these opinions are mostly formed on the basis of factors that can in principle be objectified. One example of this is the accessibility of certain amenities, such as the nearest supermarket, the nearest public transport stops and the nearest park. This accessibility or centrality of a location is analyzed by real estate experts such as brokers, real estate developers and investment managers and combined with other location-related objectifiable information – e.g., population structure, employment structure, etc. – in order to make a final judgment about the quality of a particular location.

The use of these location-specific factors that can in principle be objectified, raises the question of whether this process can be standardized and automated by means of computer-based models. Such models have been used in real estate valuation, for example, for several decades. They are usually are referred to as Automated Valuation Models (AVM) and are the subject of numerous scientific studies (see, e.g., Glumac & Des Rosiers, 2021). However, in contrast to property valuations, the process of evaluating the quality of a location is much more unstructured, requiring the linking of complex interactions and multilayered non-linear relationships. Modern machine learning models have the advantage that they can automatically capture such complex structures and thus enable their measurement, whereas this requires much more manual effort in traditional econometric models and is simply not feasible in practice. Deep neural networks, for example, are modeled on the functioning of the human brain and can independently generate insights that were previously difficult to capture with parametric and semi-parametric econometric models. In addition to neural networks, complex tree-based models, such as the XGBoost algorithm, have repeatedly shown that they are capable of delivering more accurate results than more restrictive models (see, e.g., Sangani et al., 2017; Mayer et al., 2019; Pace & Hayunga, 2020)

To the best of our knowledge, there is no approach to date that leverages the capabilities of modern machine learning algorithms to capture the quality of a real estate location in an automated manner. The objective of this paper is therefore to present a new methodology based on a state-of-the-art machine learning model and a post-hoc model agnostic[22] explainable artificial intelligence (XAI) approach, namely Shapley Additive Explanations, to evaluate real estate locations. Based on the assumption that the quality of a property's location is reflected in the individual willingness to pay for the property, and that the quality can be measured by means of objectifiable factors, this paper introduces a new approach that enables the rating of individual property locations. We call this approach "SHAP location score" (SHAP-LS). The approach is characterized by its high degree of flexibility and can be implemented in a model-agnostic manner for any machine learning algorithm and for any feature set.

Throughout the paper, we first present the theoretical foundation of the SHAP-LS and then introduce the approach in more detail. Finally, using a dataset of 26,860 residential rental listings for the city of London from 2020, and a comprehensive set of location-specific points-of-interests (POIs), we show how an empirical implementation of the SHAP-LS is possible and what results can be obtained. The results reveal that the SHAP-LS is able to identify the individually measured quality of a location and also break it down into different categories. The approach thus enables a large number of real estate players to accelerate their analyses and, in particular, to conduct them in an empirical and data-driven manner.

4.3 Theoretical Foundation

The SHAP-LS is based on two different fundamental assumptions:

I. The quality of a certain location is reflected in the price of a property through the individual willingness to pay of individual market participants and their competition with each other.

II. The quality of a particular location can be measured by means of individual features describing the location.

Assumption one represents an area that has already been much researched. Various studies show that market participants have different levels of willingness to pay with regard to different locations. A recent example can be found in Gabe et al. (2021), who examine consumer willingness to pay for location efficiency, which represents a normative component of new urbanism and describes a location based on the following five dimensions: urban design, density, land-use diversity, access to transit, and destination accessibility. Their results show that renters are willing to pay for multiple attributes of more efficient locations and require discounts for less desirable attributes. Other examples can be found in Bartholomew & Ewing (2011), Seo et al. (2014), Freybote et al. (2015) and Jauregui et al. (2019). It is therefore evident that, in addition to the property itself, location-

[22] This term describes the fact that this technique is applied after the actual training of an algorithm (= post-hoc) and can be applied for different algorithms (= model-agnostic).

related attributes also play an important role and thus decisively determine the quality of a location.

In order to deal with the second assumption, a theory or methodology is required that allows measuring the marginal price effects of individual location-describing factors. Such a theoretical foundation can be found in hedonic pricing theory. Similar to the basic assumptions mentioned above, hedonic price theory assumes that the value or price of an economic heterogenous good can be decomposed and determined on the basis of the sum of the marginally observable prices of the individual components of the good. One of the first applications of the hedonic pricing theory can be found in Court (1939) who used the theory to determine automobile prices. The first adaptations and implementations within a real estate context can be found in Lancaster (1966) and Rosen (1974), who assume that consumers derive value from different housing characteristics and that this value can be priced. Regarding the consumption of housing, consumers maximize their utility within their budget constraint. As Sirmans et al. (2005) show, the hedonic model generally takes the form:

$$Price = f(Physical\ Characteristics, Other\ Factors) \tag{22}$$

It is therefore generally assumed that, in addition to the physical characteristics of a property, such as the number of rooms or of bathrooms, other factors also play a price-determining role. These factors are generally referred to as location and neighborhood variables (Can, 1992 and Stamou et al., 2017). Location variables define the geographic location of a property, while neighborhood variables describe its social and economic environment. With regard to all three groups of variables, there is vast scientific evidence on the effect of individual variables on the price of a property. For physical aspects, for example, the studies by Sirmans et al. (2005), Kestens et al. (2006), Randeniya et al. (2017) and Metzner & Kindt (2018) provide a good overview. Location-specific aspects, on the other hand, are the focus of Hoen & Atkinson-Palombo (2016), Dumm et al. (2016), W. Seo (2018) and Turner & Seo (2021). Accordingly, the applicability of the hedonic pricing theory can be seen as well proven and suitable for locational aspects in the real estate context. Thus, the theoretical framework of our approach is based on the principles of hedonic price theory. We make use of this theory and extract the location-specific effects of a real estate price and relate them to the quality of the respective location.

Whilst the majority of empirical papers use parametric or semi-parametric models to determine marginal prices in hedonic price studies, non-parametric models are not used that often, mainly due to their lack of intrinsic explanatory power. One category of models that falls into this group are the so-called modern machine learning approaches, such as neural networks or decision-tree based models. As Chun Lin & Mohan (2011), Kok et al. (2017), Mayer et al. (2019) and Stang et al. (2022) show, these models provide a high degree of accuracy, but the results obtained cannot be explained by the model itself. However, by using so-called model-agnostic approaches, post-hoc explainability can be created (Krämer et al., 2023; Lorenz et al., 2022). Model-agnostic approaches describe a set of different new methods, which can in principle be applied to any machine learning algorithm and allow an investigation of the relationships learned by the model (Molnar, 2020).

We make use of these methods and, based on them, present a new methodology called SHAP-LS, in order to determine the quality of a location for real estate purposes. Thus, the novelty and contribution to the literature of the SHAP-LS lies essentially in the extension of traditional urban models by means of modern machine learning and model-agnostic methods. Compared to the previous parametric and semi-parametric approaches, the SHAP-LS enables taking complex and non-linear relationships into account. Especially with regard to the assessment of the quality of a location, this creates advantages, since location structures are often characterized by a multitude of different interactions (Gabe et al., 2021). With regard to the SHAP-LS, we decided to use the local model-agnostic method of Shapley Additive Explanations. Shapley values allow us to examine how much individual features contribute to the difference between individual and average prediction. This in turn allows us to capture the marginal effects of the location-determining variables at the level of individual observations and thus to measure the influence of the location quality for any given location within a dataset.

4.4 Methodology

Following our theoretical foundation, the aim of this paper is to combine hedonic pricing theory and modern machine learning algorithms to develop a purely data-driven approach for assessing the quality of a property location. The approach is designed with special attention to the necessary flexibility, adaptability, simplicity and easy implementation. The mentioned criteria are therefore important to fulfill, because only this can guarantee that the approach can be adapted for a multitude of different purposes and thus be integrated into general scientific and practical use.

As the SHAP-LS is based on a post-hoc XAI methodology, initially the training of an appropriate valuation algorithm is required. To ensure the applicability of the SHAP-LS methodology, several points regarding the training need to be considered. The first step is the selection of a dataset suitable for analysis. In addition to the price of the property[23] and its physical characteristics, especially a comprehensive set of location-related features is required. Furthermore, the dataset has to be as spatially and temporally dense as possible. Since the quality of individual locations can change over time, the temporal density prevents possible bias over the years. The spatial density allows the presentation of granular results, which are visually easier to show and additionally allow extrapolation and interpolation of the results. Due to the model agnostic character of the SHAP-LS methodology, in principle any machine learning methodology can be applied to train the underlying valuation algorithm. Given their performance characteristics, generally, the use of modern tree-based models, such as the extreme Gradient Boosting algorithm, or the use of neural networks (Antipov & Pokryshevskaya, 2012; McCluskey et al., 2012; Yoo et al., 2012; Yilmazer & Kocaman, 2020 and Ho et al., 2021) is recommended.

Post-hoc to the model training, the Shapley Additive Explanations have to be calculated on the basis of the trained machine learning algorithm. Developed by Lundberg & Lee (2017), Shapley Additive Explanations (SHAP) is a local post-hoc model-agnostic technique based on coalitional game theory in order to detect the contribution of a feature on a single prediction compared to the average prediction. One main advantage of this method is that it can be applied to all machine learning models. The SHAP value of a feature value is its contribution to the payout, which is weighted and summed over all possible feature value combinations. In a game theoretical context, the SHAP value of an observation i and feature j can be computed as follows:

$$\phi_{i,j} = \sum_{S \subseteq \{x_{i,1},\dots,x_{i,p}\}\{x_{i,j}\}} \frac{|S|!\,(p-|S|-1)!}{p!} (f\left(S \cup \{x_{i,j}\} - f(S)\right), \tag{23}$$

where $\phi_{i,j}$ denotes the features contribution, S is a coalition, $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{p}\times\boldsymbol{n}}$ is the full feature space containing p variables and n observations. Furthermore, $f(S)$ represents the prediction of a model f on S. As a result, the SHAP methodology shows a vector of the individual SHAP values for all features. It is important to note that the values are not valid on an aggregated level, but are determined separately for each individual observation in the dataset.[24] SHAP values show how the individual features influence the prediction made by the machine learning model in comparison to the average prediction. A detailed example of the SHAP values can be found in Appendix I.

[23] In the context of the SHAP-LS methodology, it is in principle possible to use both purchase or rental prices. Both reflect the observable willingness to pay for a property with certain characteristics and a certain location and can thus be used in this logic in an arbitrary manner.

[24] An example of the identification of aggregated results would be the Permutation Feature Importance (see, e.g., Krämer et al., 2023).

To finally determine the SHAP-LSs, the identified SHAP values are used. In non-technical terms, the SHAP-LS of a single observation i is computed by extracting the SHAP values of all location-specific features j and adding them up. In order to increase the significance of the results, the top and bottom percentile of all SHAP-LSs is removed and in a last step, the SHAP-LS is scaled on the basis of all other SHAP-LSs. Technically, this can be expressed as follows:

Assume that there is machine learning model f, a full feature space $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{p} \times \boldsymbol{n}}$ containing p variables and n observations. The SHAP values of each data point in $\boldsymbol{x}$ can be computed by using model f and are stored in the SHAP matrix $\boldsymbol{\phi} \in \mathbb{R}^{\boldsymbol{p} \times \boldsymbol{n}}$. Let $\boldsymbol{\phi_M} \in \mathbb{R}^{\boldsymbol{q} \times \boldsymbol{n}}$, $q \leq p$, be a subset of $\boldsymbol{\phi}$ containing the SHAP values of all location-specific features. For all observations, $i \in \{1, \dots, n\}$, all SHAP values $\phi_{i,j} \in \boldsymbol{\phi_M}$ are summed:

$$\phi_i = \sum_{j=1}^{q} \phi_{i,j}. \tag{24}$$

To obtain the final overall SHAP-LS, the top and bottom percentile of all ϕ_i are dropped and the SHAP-LSs are scaled between -1 and 1 in order to provide an easily interpretable scoring:

$$\phi_{scaled,i} = \frac{\phi_i - \min\limits_{i \in \{1,\dots,n\}}(\phi_i)}{\max\limits_{i \in \{1,\dots,n\}}(\phi_i) - \min\limits_{i \in \{1,\dots,n\}}(\phi_i)} \cdot 2 - 1. \tag{25}$$

$\phi_{scaled,i}$ is set to be the final SHAP-LS of observation i and can be interpreted as the relative importance of the location-specific features of observation i. The higher the value, the higher the marginal willingness to pay for the features used as location quality descriptors. The relative comparison across all available observations makes it possible to determine the quality of different locations on the basis of the prevailing marginal willingness to pay for all characteristics describing the location.

Besides analyzing the overall location quality, the SHAP-LS methodology can also be used to assess the individual features that drive the quality of a particular location. This can be done by creating categorical SHAP-LSs for individual related feature groups.[25] In order to

[25] Theoretically, the SHAP values of single features could be used for this kind of analysis. However, this is not recommended, as one is exposed to the capriciousness of the algorithms and data providers. Often, locational features such as the distance to the next bus stop and to the next subway station correlate highly. Consequently, the algorithm cannot distinguish perfectly between these correlated features, which can lead to a blurring of the individual SHAP values. Another reason that should not be neglected is the dependence on the categorization of the location characteristics of the data providers. In some cases, individual amenities overlap considerably, e.g., the classification of restaurants, pubs or bars. Combining several individual characteristics into categories can counteract this blurring. As a rule of thumb, it can be stated that the more data available, the smaller the categories that can be used.

compute these categorical SHAP-LSs, a subset of all location-specific SHAP values containing the SHAP values of the features in this category has to be selected, $\boldsymbol{\phi_{CAT}} \subseteq \boldsymbol{\phi_M}, \boldsymbol{\phi_{CAT}} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}}$, $m \leq q$. The following steps are similar to those above. First, for every observation $i \in \{1, \dots, n\}$ the SHAP values $\phi_{i,j} \in \boldsymbol{\phi_{CAT}}$ are summed:

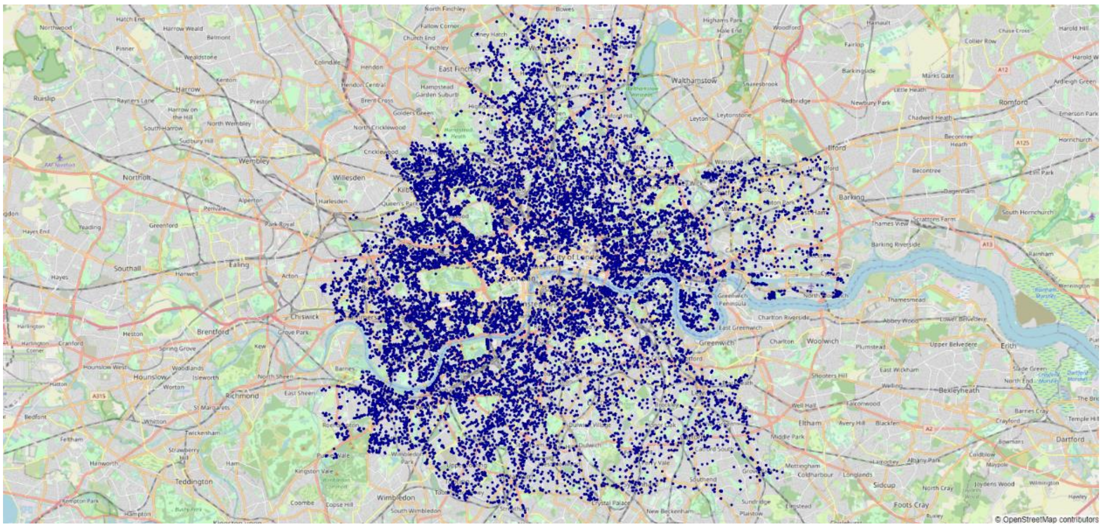
$$\phi_{CAT,i} = \sum_{j=1}^{m} \phi_{i,j}. \tag{26}$$

Again, the top and bottom percentile of all $\phi_{CAT,i}$ is excluded and all $\phi_{CAT,i}$ are scaled between -1 and 1 to ensure easy interpretation. Therefore, $\phi_{CAT,scaled,i}$ defines the final categorical SHAP-LS and can be interpreted as the relative importance of a feature category for an observation i regarding the marginal willingness to pay.

4.5 Application

The following section deals with an empirical implementation of the presented SHAP-LS. We use an exemplary dataset from the city of London for this purpose. The dataset consists of 26,860 residential rental properties that were listed on different Multiple Listing Systems (MLS) in 2020. Figure 4.1 shows how the observations are distributed across the city.

Figure 4.1: Distribution of individual observations across the city of London



Notes: This figure shows the spatial distribution of all available observations within the city of London. The 26,860 observations are distributed over the entire city, whereby the spatial density is particularly high in the center and decreases towards the outskirts. The data was generated by webscrapping from different MLSs, and cleaned using standardized procedures.

In addition to the rental price, a set of hedonic features defining the physical characteristics of the individual properties and a time variable to capture temporal and seasonal effects are available. To compute the SHAP-LS, a comprehensive set of location-specific features is enriched, using the geocoordinates of the properties. We capture the location quality by means of the distance to the nearest major amenities. The distance is calculated using Euclidean proximity. The integration of further location-related features is in principle arbitrarily representable. Table 4.1 summarizes the descriptive statistics.

Table 4.1: Descriptive statistics

Variable	Category	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Rent GPB/p.m.	Price	Metric	2111.4	1948	769.85	4546.00	346.00
Bedrooms	Physical	Integer	1.97	2.00	0.80	5.00	1.00
Month of listing	Physical	Integer	5.63	6.00	3.32	12.00	1.00
New build	Physical	Binary	0.07	0.00	0.25	1.00	0.00
Size sqm	Physical	Metric	64.68	61.04	23.44	248.70	20.07
Bus stop_dist	Location	Km	0.12	0.11	0.08	0.79	0.00
Bus station_dist	Location	Km	2.33	2.03	1.44	8.09	0.03
Railway station_dist	Location	Km	0.49	0.43	0.30	3.18	0.01
Car sharing_dist	Location	Km	0.42	0.31	0.42	4.43	0.00
Bicycle rental_dist	Location	Km	0.94	0.32	1.21	7.12	0.00
Motorway junction_dist	Location	Km	4.5	4.49	2.23	10.87	0.04
Doctors_dist	Location	Km	0.55	0.48	0.36	2.77	0.00
Hospital_dist	Location	Km	1.01	0.89	0.61	4.28	0.01
Pharmacy_dist	Location	Km	0.39	0.33	0.27	2.56	0.00
Dentist_dist	Location	Km	0.55	0.44	0.43	3.96	0.00
Optician_dist	Location	Km	0.72	0.60	0.56	4.66	0.00
School_dist	Location	Km	0.24	0.21	0.13	0.99	0.00
Kindergarten_dist	Location	Km	0.61	0.52	0.42	3.77	0.00
University_dist	Location	Km	1.44	1.19	1.08	6.73	0.01
College_dist	Location	Km	1.02	0.84	0.79	5.27	0.01
Restaurant_dist	Location	Km	0.24	0.18	0.22	2.18	0.00
Fast food_dist	Location	Km	0.3	0.24	0.22	1.79	0.00
Cafe_dist	Location	Km	0.24	0.19	0.21	2.23	0.00
Greengrocer_dist	Location	Km	0.95	0.78	0.73	6.00	0.00
Bakery_dist	Location	Km	0.62	0.51	0.49	4.28	0.00
Supermarket_dist	Location	Km	0.36	0.32	0.24	1.93	0.00
Department store_dist	Location	Km	1.39	1.19	1.00	6.73	0.01
Mall_dist	Location	Km	1.62	1.31	1.13	6.69	0.02
Clothes_dist	Location	Km	0.55	0.46	0.43	4.23	0.00
Kiosk_dist	Location	Km	1.48	1.20	1.05	6.97	0.01
Atm_dist	Location	Km	0.39	0.31	0.31	3.07	0.00
Post office_dist	Location	Km	0.48	0.44	0.27	2.35	0.00
Post box_dist	Location	Km	0.14	0.13	0.09	0.87	0.00
Hairdresser_dist	Location	Km	0.35	0.26	0.36	3.90	0.00
Laundry_dist	Location	Km	0.36	0.28	0.32	3.41	0.00
Sports centre_dist	Location	Km	0.52	0.47	0.31	2.49	0.00
Park_dist	Location	Km	0.27	0.23	0.16	1.26	0.00
Playground_dist	Location	Km	0.32	0.27	0.21	1.44	0.01
Swimming pool_dist	Location	Km	1.14	1.02	0.71	5.54	0.01
Bar_dist	Location	Km	0.53	0.41	0.45	4.10	0.00
Nightclub_dist	Location	Km	1.53	1.20	1.18	7.15	0.01
Pub_dist	Location	Km	0.22	0.18	0.17	1.71	0.00
Beer garden_dist	Location	Km	2.51	2.24	1.42	9.05	0.01
Prison_dist	Location	Km	4.15	3.81	2.36	14.08	0.09
Wastewater plant_dist	Location	Km	5.48	5.46	2.77	14.38	0.03
Graveyard_dist	Location	Km	1.43	1.37	0.72	3.99	0.04
Windmill_dist	Location	Km	3.91	3.95	1.78	10.25	0.02

Notes: This table reports the descriptive statistics of the dataset. In addition to the physical features describing the property itself and a temporal variable, distances to the most important POIs were added for all observations, so as to describe the spatial location of the individual properties. The distances were determined by means of the Euclidean distance and are given in kilometers (Km). The selection of the parameters was in accordance with other publications in the real estate literature (see, e.g., Metzner & Kindt, 2018). The parameter "Rent GPB/sqm/p.m." is the dependent variable in our model and describes the asking rent for the individual properties.

As an underlying machine learning model, the XGBoost algorithm is chosen. This is a computationally effective and highly efficient sequential tree-based ensemble learning method. For more information about tree-based methods, ensemble learning and (extreme) gradient boosting, the reader is advised to read Hastie et al. (2001) and Chen & Guestrin (2016). In order to ensure that the machine learning model provides reliable results, the model was tested by using five-fold cross validation on four different evaluation metrics, namely the Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE) and the 10/20 Percent Error Buckets (PE(10) and PE(20)). A more detailed description of the error metrics can be found in Appendix II. Table 4.2 shows the model performance results, which indicate a strong and robust performance of the XGBoost. Therefore, it is well suited for a post-hoc analysis and the SHAP-LS can be calculated.

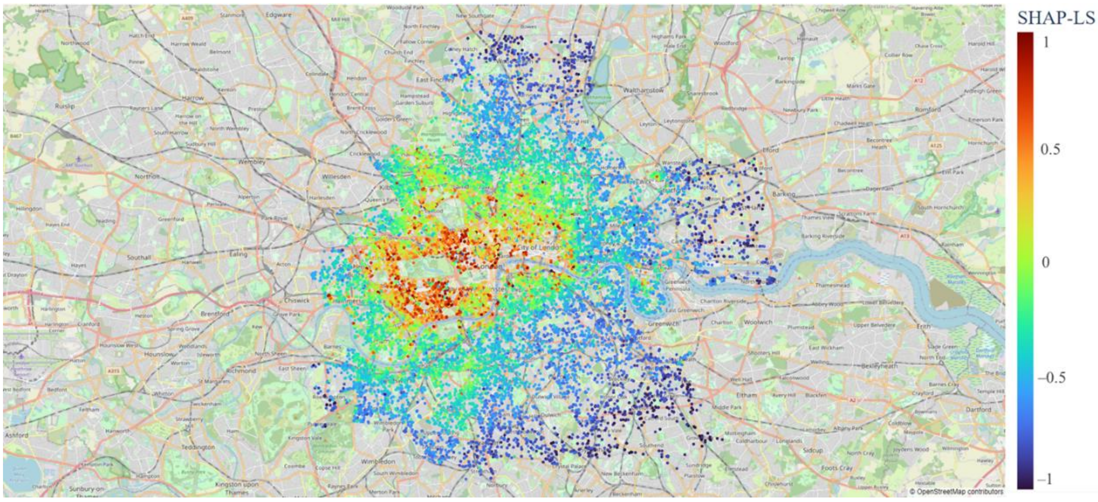
Table 4.2: Overview model performance

Metrics	XGBoost
MAPE	0.1536
MdAPE	0.0968
PE(10)	0.5124
PE(20)	0.7851

Notes: This table shows the results of the XGBoost model. A detailed description of the evaluation metrics used can be found in Appendix I. The results indicate a strong model performance. The median deviation of the XGBoost (MdAPE) is below 10%, which indicates a solid valuation result and allows further use of the results for a post-hoc analysis.

Figure 4.2 visualizes the results of the calculated SHAP-LSs. With regard to the visual representation, relatively attractive locations are shown in red, and locations seen as relatively unattractive in blue. The selected color gradient makes it easy to recognize differences between the individual subareas. As the picture shows, the SHAP-LS methodology recognizes different levels of location attractiveness or location quality within the city of London. It can be seen that particularly attractive locations are distributed around the center of London and are predominantly found in the districts of Mayfair, Kensington, South Kensington, Knightsbridge, Brompton, Chelsea and Convent Garden. In terms of the SHAP-LS methodology, this shows that in these parts of the city, the marginal willingness to pay for location-defining features is particularly high and takes up a large share of the rent paid in each case. In contrast, relatively unattractive locations are found in particular in the outer boroughs of London. This empirical result is consistent with the theoretical assumption that attractive locations are found in the central areas of a city. With reference to the city of London, it can be seen that the SHAP-LS is capable of identifying the theoretical and presumed attractiveness levels of different locations on the basis of empirical data. Since the score is always calculated at the level of individual observations, the attractiveness of a location can be determined at the block or street level, in contrast to a conventional manual analysis, which usually refers to individual neighborhoods.

Figure 4.2: Overall SHAP location score

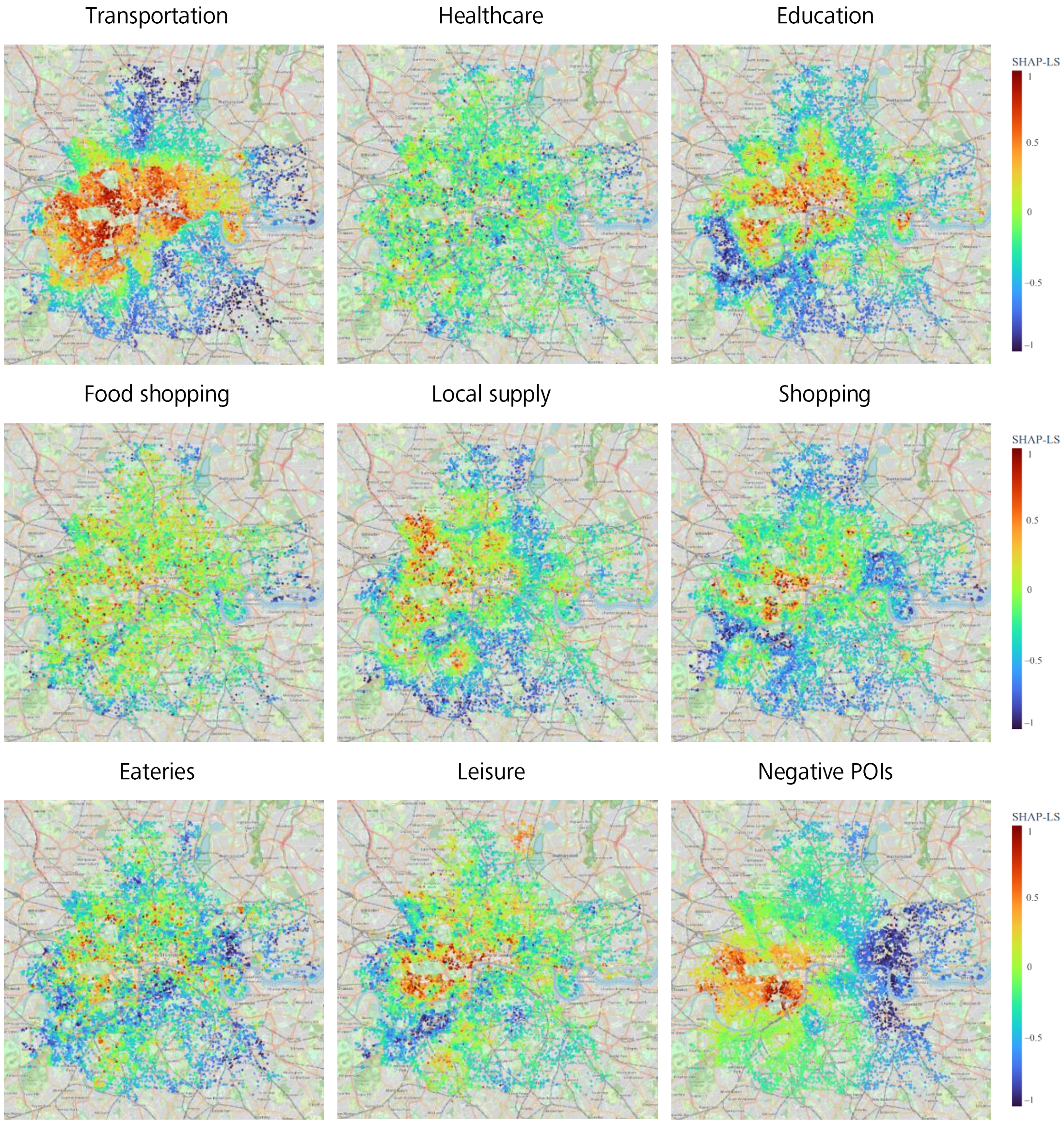


Notes: This figure visualizes the final results of the calculated SHAP-LSs. For each individual observation within the dataset, the corresponding SHAP-LS was determined using the methodology described above. The visual representation of the scores clearly shows that there are key differences with regard to the quality of different real estate locations. It is particularly noticeable that the central locations near the CBD seem to be characterized by above-average quality.

In a next step, we divide the location-specific features into different categories to provide more insight into how the different amenities and disamenities affect the individual SHAP-LSs. Overall, we divide the amenities into nine categories: Transportation, healthcare, education, food shopping, local supply, shopping, eateries, leisure and negative POIs, like the distance to the next prison or graveyard. Figure 4.3 visualizes the final results.

Focusing on the highlights of the maps, the visualization of the category “Education”, for example, shows some interesting patterns. It is clearly evident that, apart from the city center, there seem to be individual clusters within London that are characterized by a high level of payment dispersion with regard to the spatial distance to educational institutions. A more detailed analysis shows that these clusters are located close to certain educational centers (e.g., universities). The spatial proximity to higher educational institutions thus initially has a positive influence on the quality of a property location in London. With regard to the category “Transportation”, a clear pattern also emerges. The marginal price influence of POIs in this category is almost ring-shaped around the city center and is arranged according to the local public transport axes. Furthermore, it can be seen that there are also individual clusters here, which in turn are determined by local traffic centers in their midst. This picture clearly shows that a good public transport connection is of importance and represents an essential aspect for the determination of location quality in London.

Figure 4.3: Categorical SHAP location score



Notes: This figure maps the results of the calculated categorical SHAP-LSs. The results show how the individual amenities and disamenties of the individual categories influence the asking rent. The maps indicate that the effect varies by category and location. For example, there are categories whose influence seems to be relatively evenly distributed across the city of London (e.g., Healthcare) and in contrast, other categories clearly show that certain locations are of higher quality than others in terms of spatial proximity to the POIs of these categories (e.g., Transportation).

In contrast, the effect for POIs in the categories "Food Shopping" and "Healthcare", for example, is very small and the marginal willingness to pay is almost the same across the city. Only at the outermost edges does this effect reverse. For both categories, this indicates either that these types of POIs are available equally throughout the city or that they do not play an important role in the individual determination of willingness to pay. Due to the frequency of necessary food shopping and the generally good supply within the city, the assumption for the category "Food Shopping" lies with the former argumentation. For the category "Healthcare", on the other hand, the presumption lies with the second argument, since these typically do not determine a person's everyday life and thus initially appear less important. Last but not least, it is also worth taking a look at the category "Negative POIs". Here, an east-west gradient is particularly noticeable, which is not the case in any of the other categories. It can be seen that there is a greater willingness to pay for a particular location, especially in the city center, as this is influenced only minimally by negative POIs. This shows that it is not only important to analyze positive aspects to determine the quality of a location, but also to include points that are perceived as negative. The selected examples illustrate well how the categorical SHAP-LSs can be used to explain the results of the overall SHAP-LS and to perform more detailed analyses.

4.6 Conclusion

While Automated valuation models (AVM) have become popular in the field of real estate valuation, there are hardly any applications for the valuation of the quality of a real estate location. Accordingly, the objective of our work is to apply the principles of AVMs to the field of location analytics for real estate purposes. For this purpose, we have developed a new approach, named SHAP location score (SHAP-LS), which is able to assess different real estate locations on the basis of their quality. The approach results in an overall score that indicates whether the quality of a particular location is strong or weak. In addition to the overall score, the SHAP-LS method also allows us to disaggregate the score into different categories and thus analyze which factors have a greater influence on the quality of the location and which play only a subordinate role. The SHAP-LS approach is based on two basic assumptions. The first implies that the quality of a certain location is reflected in the price of a property by means of the individual willingness to pay of individual market participants and their competition with each other. The second assumption proposes that the individual differences in quality can be measured by objectifiable factors. These two basic assumptions essentially allow the theoretical framework of hedonic price theory to be applied. The combination of a non-parametric machine learning algorithm and the local post-hoc model-agnostic explainable artificial intelligence (XAI) method of Shapley Additive Explanations allows us for the first time to exploit the properties of modern machine learning methods in this particular domain, and thus to capture non-linearities and higher-order interactions, which appropriately play an important role in location analytics. Previously this was only possible to a limited extent using parametric or semi-parametric methods.

An application of our methodology to a dataset of the city of London shows that the SHAP-LS method is able to extract the value - and thus also location quality - determining factors of real estate prices on a granular level and thus to show where within the city the quality of a location is particularly good and where it is not. The results show that there are significant differences with regard to the individual willingness to pay for different levels of location quality and that this willingness seems to be highest in the central locations of the city. While this result seems to be nothing new in principle, this time an analysis can be made not only at the level of districts, as is usually the case, but also at the street level. A presentation of the results according to different categories also allows us to further analyze which group of points of interest contributes how strongly to the location quality. The results obtained show that the degree of willingness to pay for location quality essentially depends on the spatial proximity to public transport stops, educational institutions, shopping facilities and negative POIs. Other frequently used categories, such as proximity to food shopping POIs, play a lesser role.

The SHAP-LS represents an initial approach to an automated valuation of real estate locations, which is suitable for a variety of different real estate players. The approach can be applied post-hoc to any common machine learning method and can be flexibly adapted to the respective needs. The approach is not limited to certain features, and can be adapted to the existing data structure by the modeler. Regarding academia, the presented approach extends traditional urban models by the use of modern non-parametric valuation algorithms and thus allows an extended consideration of economic and econometric phenomena. For practitioners, the approach enables quickly obtain an empirical assessment of the quality of different locations within a city. Thus, the transparency of local markets can be increased, and more well-founded or more objectified decisions can be made.

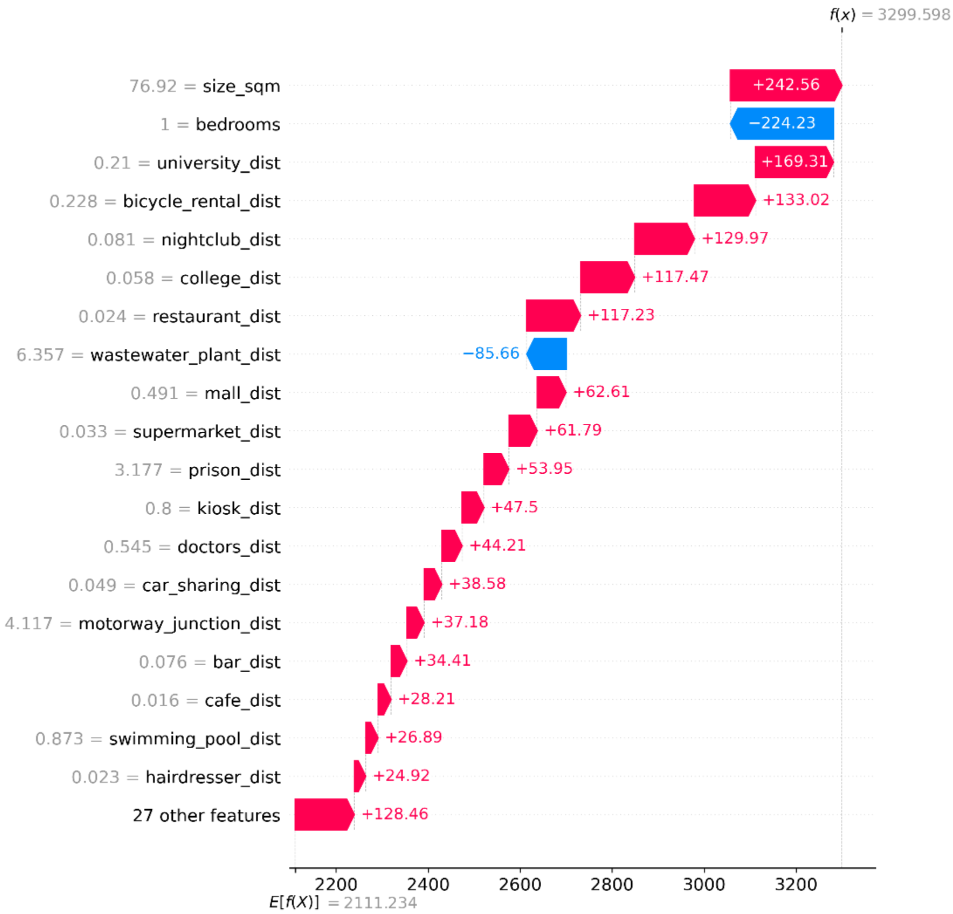
While the initial intention of this paper was to introduce the new SHAP-LS, there are already other use cases that are worth investigating in terms of future work. For example, the application of the approach to different asset classes needs to be examined and investigated. Furthermore, the use of the methodology to investigate the quality of a location over time seems very promising. Furthermore, an in-depth examination of the respective effect of potential negative externalities, such as road noise or air pollution, could provide valuable insights. Accordingly, we encourage other authors to explore further and yet unknown areas on the basis of the SHAP-LS methodology.

4.7 Appendix

4.7.1 Appendix I – SHAP Value Example

Figure 4.4 shows an example of the SHAP values of the model and dataset used in this paper i.e., a rent forecast for the city of London. On the y-axis, the features are ranked starting with the feature that contributes least (bottom left) to the individual estimate compared to the average rent price prediction up to the feature that represents the largest differentiating factor (up right). On the x-axis, the predicted rental price is shown. E[f(x)] denotes the price of the average prediction, £2,111, and f(x) the prediction of the selected observation, £3,300. Red arrows indicate a positive, and blue arrows a negative contribution of the individual feature. In this example, the apartment size of 77 square meters has a positive impact of £243, compared to the average prediction, and the fact that the apartment has only one bedroom has a negative impact of about -£224.

Figure 4.4: SHAP value example



Notes: This figure shows an example of how SHAP values work. The example is an actual observation from the London dataset used for this paper. The plot shows that the average prediction of the XGBoost across all observations is £2,111, and for the observation shown here, £3,300. How the difference between the average prediction and this observation is composed can be easily seen thanks to the waterfall-like structure of the SHAP values.

4.7.2 Appendix II – Evaluation Metrics

Table 4.3: Evaluation metrics

Error	Formula	Description
Mean Absolute Percentage Error (MAPE)	$MAPE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|$	Mean of all absolute percentage errors. A lower MAPE signals higher overall prediction accuracy in percent.
Median Absolute Percentage Error (MdAPE)	$MdAPE(y, \hat{y}) = median(\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|)$	Median of all absolute percentage errors. A lower MdAPE denotes greater precision in percent without being sensitive to outliers.
Error buckets (PE(x))	$PE(x) = 100\left\|\frac{y_i - \hat{y}_i}{y_i}\right\| < x$	Percentage of predictions where the relative deviation is less than x%, with x being 10 and 20. A larger PE(x) signals a lower variation in the predictions.

Notes: This table reports the evaluation metrics used to determine the valuation accuracy of the XGBoost and OLS algorithm. All three metrics are regularly used to assess the quality of different valuation algorithms. The choice of several metrics in total, allows a more differentiated statement to be made than would be the case with just one metric.

4.8 References

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.

Bartholomew, K., & Ewing, R. (2011). Hedonic Price Effects of Pedestrian- and Transit-Oriented Development. *Journal of Planning Literature*, 26(1), 18–34.

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

Chen, T., & Guestrin, C. (2016). XGBoost.

Chun Lin, C., & Mohan, S. B. (2011). Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis*, 4(3), 224–243.

Court, A. T. (1939). *Hedonic price indexes with automotive examples*. In The Dynamics of Automobile Demand, General Motors Corporation, New York, 99–117.

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.

Freybote, J., Sun, H., & Yang, X. (2015). The Impact of LEED Neighborhood Certification on Condo Prices. *Real Estate Economics*, 43(3), 586–608.

Gabe, J., Robinson, S., & Sanderford, A. (2021). Willingness to Pay for Attributes of Location Efficiency*. The Journal of Real Estate Finance and Economics*, 1–35.

Glumac, B., & Des Rosiers, F. (2021). Practice briefing – Automated valuation models (AVMs): Their role, their advantages and their limitations. *Journal of Property Investment & Finance*, 39(5), 481–491.

Hastie, T., Friedman, J. & Tibshirani, R. (2001). *The Elements of Statistical Learning.* New York: 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: Astudy of Home Value Impacts in Densely Populated Massachusetts. *Journal of Real Estate Research*, 38(4), 473–504.

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.

Kestens, Y., Thériault, M., & Des Rosiers, F. (2006). Heterogeneity in hedonic modelling of house prices: Looking at buyers' household profiles. *Journal of Geographical Systems*, 8(1), 61–96.

Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal of Portfolio Management*, 43(6), 202–211.

Krämer, B., Stang, M., Nagl, C., & Schäfers, W. (2023). Explainable AI in a Real Estate Context - Exploring the Determinants of Residential Real Estate Values. *Journal of Housing Research*, Advanced online publication.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.

Lorenz, F., Willwersch, J., Cajias, M., & Fuerst, F. (2022). Interpretable machine learning for real estate market analysis. Journal of *Real Estate Economics*, Forthcoming.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Proc. *Advances in Neural Information Processing Systems*, 30, 4768–4777.

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., Davis, P., Haran, M., McCord, M., & McIlhatton, D. (2012). The potential of artificial neural networks in mass appraisal: The case revisited. *Journal of Financial Management of Property and Construction*, 17(3), 274–292.

Metzner, S., & Kindt, A. (2018). Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties. *International Journal of Housing Markets and Analysis*, 11(1), 73–100.

Molnar, C. (2020). Interpretable Machine Learning. Lulu.com.

Pace, R. K., & Hayunga, D. (2020). Examining the Information Content of Residuals from Hedonic and Spatial Models Using Trees and Forests. *The Journal of Real Estate Finance and Economics*, 60(1-2), 170–180.

Parker, D. (2016). *International Valuation Standards: A Guide to the Valuation of Real Property Assets*. Chichester: Wiley Blackwell.

Randeniya, T. D., Ranasinghe, G., & Amarawickrama, S. (2017). A model to Estimate the Implicit Values of Housing Attributes by Applying the Hedonic Pricing Method. *International Journal of Built Environment and Sustainability*, 4(2), 113–120.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Sangani, D., Erickson, K. & Hasan, M.A. (2017). *Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting*. In IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Orlando.

Seo, K., Golub, A., & Kuby, M. (2014). Combined impacts of highways and light rail transit on residential property values: a spatial hedonic price model for Phoenix, Arizona. *Journal of Transport Geography*, 41, 53–62.

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.

Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, 13(1), 1–44.

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

Stang, M., Krämer, B., Nagl, C., & Schäfers, W. (2022). From human business to machine learning—Methods for automating real estate appraisals and their practical implications. *Zeitschrift Für Immobilienökonomie*, Advanced online publication.

Turner, T. M., & Seo, Y. (2021). House Prices, Open Space, and Household Characteristics. *Journal of Real Estate Research*, 43(2), 204–225.

Yilmazer, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy*, 99, 104889.

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.

5 Conclusion

5.1 Executive Summary

This part of the thesis briefly summarizes the content of all three papers. It discusses the objectives of each study, data and methodologies used, as well as results and implications for science and practice.

Paper 1: From Human Business to Machine Learning – Methods for Automating Real Estate Appraisals and their Practical Implications

---

Problems and Objective

Various ways are known in science and practice to apply the sales comparison approach in the context of Automated Valuation Models (see, Isakson, 2002). Besides the integration of filters and similarity functions, well-established hedonic price models and modern machine learning approaches can also be used (see, e.g., Pagourtzi et al., 2003 and Bogin & Shui, 2020). Currently, the use of AVMs in the real estate lending process is only allowed for supporting purposes in most countries and not as a value-determining tool (Matysiak, 2017 and Downie & Robson, 2008). Although there are now regulatory efforts to include AVMs in the lending process, this is only possible if the traceability, auditability, robustness and resilience of the inputs and outputs can be guaranteed (European Banking Authority, 2020). So far, it remains unclear which of the abovementioned methods meet these requirements. While there is an ongoing debate about allowing the use of AVMs based on already established methods such as similarity functions or OLS regressions within the lending process, the application of modern machine learning methods is almost completely absent from the regulatory discussion. This is in fact due to the "black box" label of modern machine learning techniques. However, in recent years, there have been various approaches to opening this black box; see for example by Friedman (2001), Goldstein et al. (2015), Lundberg & Lee (2017) and Apley & Zhu (2020). Through these approaches, the requirements of the supervisory authority for tractability and audibility can be considered.

Therefore, the question arises as to whether modern machine learning algorithms should also be considered by the regulatory body. The objective of this paper is to contribute to this ongoing debate and deliver further insights, based on a unique nationwide dataset, into the optimal use of modern machine learning algorithms for AVMs from a theoretical and practical point of view.

Methodology and Data

For the purpose of this paper, an automation of the sales comparison method by using filters and similarity functions, referred to as Expert Function (EXF), two hedonic price functions based on Ordinary Least Squares (OLS) and Generalized Additive Models (GAM), as well as the machine learning approach eXtreme Gradient Boost (XGBoost), are compared with each other.

The analysis is based on a data set of 1,212,546 residential properties across Germany. The data set is provided by a large German banking group and originates from valuations of standard residential real estate lending. The data was collected between 2014 and 2020. The market value of the properties is used as the dependent variable, and the available physical characteristics of each property are used to describe their condition and features as accurately as possible. All properties are georeferenced, making it possible to add a spatial gravity layer in order to account for spatial information. Features describing the location and neighborhood of the observations are added via Open Street Map and Acxiom.

Results and their Contribution to Science and Practice

In particular, this paper answers the question of whether more thought should be given to the future use of machine learning algorithms in the context of AVMs. As the results show, the machine learning approach XGBoost achieves the highest overall accuracy in the valuation of standard residential properties in Germany. Furthermore, the results of this paper show that for designing an AVM, there is no “one size fits all”. Although the XGBoost is the best performer across the country, there are also parts in Germany where the EXF, OLS, or GAM are best suited for estimating market values. In this context, it is particularly evident that the respective data availability seems to play a role. In summary, the study provides evidence that the use of machine learning algorithms for AVMs is beneficial in many situations and therefore, their approval should indeed be discussed by the regulatory authorities.

Although AVMs represent a wide field in the literature, this paper is the first to compare a filter- and similarity-based AVM approach, two well-established hedonic methods and a modern machine learning approach on a nation-wide level. The results provide important insights into the practical application of AVMs and the discussion as to whether the usage of machine learning algorithms for the lending process should be allowed from a regulatory perspective. Accordingly, the paper contains both important theoretical and practical implications that will help to further optimize the use of AVMs in the field of real estate valuation and increase their acceptance.

Paper 2: Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values

Problems and Objective

Understanding real estate markets and its drivers is arguably one of the most important areas of real estate research. Compared to other asset classes, real estate is a heterogeneous asset and properties differ from one another in terms of their features. There is a large body of literature dealing with the factors which have a significant influence on the value or price of a property, subsumed under the term Hedonic Price Models. They are usually based on parametric and semi-parametric methods like the Ordinary Least Square approach (see, e.g., Malpezzi, 2002; Sirmans et al., 2005; Schulz et al., 2014) or the Generalized Additive Models (see, e.g., Bourassa et al., 2007; Bourassa et al., 2010; Brunauer et al., 2010). In recent years, more advanced statistical and modern machine learning methods have attracted interest in the real estate community, as they are often less restrictive in terms of their model structure and thus better able to capture complex relationships. However, machine learning applications are usually criticized for their lack of transparency and are therefore often referred to as "black boxes" (see, e.g., Din et al., 2001; McCluskey et al., 2013). To overcome this problem, so-called eXplainable Artificial Intelligence (XAI) approaches have been developed. These approaches use model-agnostic frameworks to reveal the modes of operations of machine learning algorithms and thus help to make their mode of action more transparent.

In a real estate context, so far, XAI approaches have been explored only to a limited extent, therefore the objective of this paper is to show how XAI methods can be used to make the deep hidden patterns of residential real estate markets interpretable for human beings. Furthermore, the previous literature has mainly focused on the identification of non-linearities, but falls short on the interpretation of their economic implications. Accordingly, this paper not only focuses on identifying reliable and unbiased relations between features and residential property prices, but also discusses their economic implications.

Methodology and Data

To conduct the study, the modern machine learning algorithm XGBoost is first used to identify the individual linear and non-linear relationships. Then, two XAI methods are employed to make the learned relationships visible. First, Permutation Feature Importance (PFI), first introduced by Breiman (2001), is used to analyze which features actually influence the value of a property. Next, so-called Accumulated Effects Plots (ALE), established by Apley & Zhu (2020), are applied to further make a statement about the effects themselves and whether non-linear relationships can be identified or not. In particular, the former is used as a basis for the latter, to identify which variables have the greatest impact on property values.

The paper uses a dataset of 81,166 residential properties for the seven largest cities of Germany. The data originate from the years 2014 to 2020. To detect differences between different subtypes of residential properties, the dataset is further split into two groups. The first consists of 61,763 condominiums and the second of 19,403 single-family homes. The dataset is provided by a large German banking group and originates from their valuation department. The market value per square meter of the properties is used as the target variable. In addition to the dependent variable, a set of features defining the structural characteristics of the properties is used. All properties are georeferenced, making it possible to add a spatial gravity layer to account for spatial information. Features describing the location and neighborhood of the properties are added via Open Street Map and Acxiom.

Results and their Contribution to Science and Practice

The results of the paper reveal that the same value-determining features play an important role for both condominiums and single-family homes. However, there are fundamental differences within the two property types with regard to the shape of the individual ALE plots and thus the influence of the respective feature on the market value of a property. Furthermore, non-linear relationships are identified for the majority of features. Generalized rules of thumb such as "the larger the living area, the lower the market value per square meter" are refuted by the findings for condominiums, but can be confirmed for single-family homes.

In summary, the results show how important it is for both real estate research and practice to conduct data-driven analyses with the help of modern machine learning and XAI approaches, in order to gain important market insights and, if necessary, to update long-established assumptions regarding the determinants of real estate market values. This is especially important for mortgage underwriters, valuation firms and regulatory authorities and, thus, of considerable interest to most of the real estate community.

Paper 3: Changing the Location Game – Improving Location Analytics with the Help of Explainable AI

Problems and Objective

Besides its structural and economic characteristics, the location of a property is probably one of the most important determinants of its underlying value. While there have been standardized and globally accepted methods for assessing property values (see, e.g., Parker, 2016), the assessment of the location quality at a given site remains a kind of arbitrary estimation based on more or less subjective individual opinions. Interestingly, these opinions are mostly formed on the basis of factors that can in principle be objectified. One example of this is the accessibility of certain amenities, such as the nearest supermarket, the nearest public transport stops and the nearest park. This raises the question of whether this process can be standardized and automated by means of computer-based models.

So far, there is no approach to date that leverages the capabilities of modern machine learning algorithms to capture the quality of a real estate location in an automated manner. The objective of this paper is therefore to present a new methodology based on a state-of-the-art machine learning model and a post-hoc model agnostic explainable artificial intelligence (XAI) approach, namely Shapley Additive Explanations, to evaluate real estate locations.

Methodology and Data

This paper introduces a new approach that enables the rating of individual property locations. The approach is called "SHAP location score" (SHAP-LS) and is based on a post-hoc XAI methodology known as Shapley Additive Explanations (developed by Lundberg & Lee (2017)). The approach combines hedonic pricing theory and modern machine learning algorithms to develop a purely data-driven approach for assessing the quality of a property location. The approach is characterized by its high degree of flexibility and can be implemented in a model-agnostic manner for any machine learning algorithm and for any feature set.

For an exemplary empirical implementation of the presented SHAP-LS a dataset from the city of London is used. The dataset consists of 26,860 residential rental properties that were listed on different Multiple Listing Systems (MLS) in 2020. In addition to the rental price, a set of hedonic features defining the physical characteristics of the individual properties and a time variable to capture temporal and seasonal effects are available. To compute the SHAP-LS, a comprehensive set of location-specific features is enriched, using the geocoordinates of the properties.

Results and their Contribution to Science and Practice

The SHAP-LS represents an initial approach to an automated valuation of real estate locations, which is suitable for a variety of different real estate players. The approach can be applied post-hoc to any common machine learning method and can be flexibly adapted to the respective needs. The empirical application of the SHAP-LS methodology to a dataset of the city of London shows that the approach is able to extract the value - and thus also location quality - determining factors of real estate prices on a granular level and thus to show where within the city the quality of a location is particularly good and where it is not. The results show that there are significant differences with regard to the individual willingness to pay for different levels of location quality and that this willingness seems to be highest in the central locations of the city.

Regarding academia, the presented approach extends traditional urban models by the use of modern non-parametric valuation algorithms and thus allows an extended consideration of economic and econometric phenomena. For practitioners, the approach enables a quick and straightforward empirical assessment of the quality of different locations within a city. Thus, the transparency of local markets can be increased, and more well-founded or more objectified decisions can be made.

5.2 Final Remarks

The appraisal industry is currently undergoing a significant transformation process, triggered by a changing regulatory environment, growing industry-specific challenges, and the general influence of advancing digitalization. This fundamentally changes how the process of assessing a property is conducted. Back in 2015, the Royal Institution of Chartered Surveyors stated that new technologies, especially the emergence of Artificial Intelligence and machine learning algorithms, would fundamentally change the work of appraisers (Royal Institution of Chartered Surveyors, 2015). Accordingly, their profession is facing a dynamic and evolving future. The fear that real estate assessors will be replaced by AI and computers in the future is extensively discussed (see, e.g., Ko & Shin, 2021); however, it is not expected that this will actually happen in the near future. Rather, the appraisal profession will make use of AI technologies to overcome industry-specific challenges, such as the global shortage of skilled workers.

So, it is not AI that is going to do the job of an appraiser in the future, rather an appraiser that knows how to use AI is going to do the job. Modern machine learning algorithms will become important tools that help real estate evaluators to manage their professional work and will make their profession future proof. All of this means that the role of appraisers is currently changing and will keep on changing significantly in the coming years. In general, their role will be more focused on data handling and information processing than traditional evaluation tasks. The pool of skilled real estate assessors will be used in areas where they are most needed, and the rest of the work will be more automated. For example, instead of searching for comparable properties and other manual tasks, an appraiser is going to be more focused on setting standards for data collection, data cleaning, and algorithm designing. However, this requires a fundamental understanding of how such new technologies work and how they can be used efficiently.

This dissertation presents a significant contribution to further enhance this understanding. The three individual papers show how modern machine learning algorithms can be optimally applied for real estate appraisal purposes. Specifically, they show how these algorithms can be used for the real estate valuation task with a focus on residential properties, to analyze individual real estate markets while overcoming the "black box" image of these algorithms, and to determine the quality of the location of a property in a data-driven and automated way. All three contributions answer different questions that provide new insights from both scientific and practical perspectives. The findings show that the use of modern machine learning methods delivers promising results and that they are able to solve tasks in a more efficient and, therefore, cost-effective way. In summary, this dissertation provides a good overview of what is possible with modern machine learning algorithms and how their use can decisively support the necessary efficiency increase in real estate appraisal. This provides an important and decisive contribution to the further development and future use of AI methods to fully leverage their potential in real estate appraisal.

While the focus of this dissertation is on the area of residential properties, the methodologies employed are equally applicable to the area of commercial properties. Regarding further research potential, this area presents promising research opportunities. Looking at previous publications, it becomes evident that there have been only a few contributions that have delved into the subject matter in depth (see, e.g., Deppner et al., 2023). Often, the availability of adequate data impedes its corresponding application. However, this aspect is currently changing due to the increasing awareness of the importance of real estate-related data and its increased collection (see, e.g., Royal Institution of Chartered Surveyors, 2017). Moreover, the use of modern machine learning algorithms promises interesting results regarding the further exploration of specific sustainability aspects. First studies on this topic are already available (see, e.g., Mohd et al., 2020). Nevertheless, the need is still not fully satisfied and will continue to rise due to the increasing importance of the sustainability topic within the real estate industry.

Therefore, the world of real estate remains an exciting research area with a promising future. These are inspiring times, and it is worth looking forward to and actively contributing to the further transformation of the industry.

5.3 References

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.

Bogin, A.N. & Shui, J. (2020). Appraisal Accuracy and Automated Valuation Models in Rural Areas. *The Journal of Real Estate Finance and Economics*, 60(1-2), 40–52.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial Dependence, Housing Submarkets, and House Price Prediction. *The Journal of Real Estate Finance and Economics*, *35*(2), 143–160.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2010). Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions. *Journal of Real Estate Research*, *32*(2), 139–159.

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

Brunauer, W. A., Lang, S., Wechselberger, P., & Bienert, S. (2010). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *The Journal of Real Estate Finance and Economics*, *41*(4), 390–411.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental Variables and Real Estate Prices. *Urban Studies*, *38*(11), 1989–2000.

Deppner, J., von Ahlefeldt-Dehn, B., Beracha, E. & Schaefers, W. (2023). Boosting the Accuracy of Commercial Real Estate Appraisals: An Interpretable Machine Learning Approach. *The Journal of Real Estate Finance and Economics*, Advanced online publication.

Downie, M.-L. & Robson, G. (2008). *Automated Valuation Models: an International Perspective*. In RICS Automated Valuation Models Conference: AVMs Today and Tomorrow, London.

European Banking Authority (2020). *Final Report: Guidelines on loan origination and monitoring.* European Banking Authority, Paris.

Friedman, J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189–1232.

Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*, 24(1), 44–65.

Isakson, H. (2002). The Linear Algebra of the Sales Comparison Approach. *Journal of Real Estate Research*, 24(2), 117–128.

Ko, H. & Shin, J. (2021). A Study on the Use and Risk of Artificial Intelligence in the Appraisal Industry. *Ilkogretim Online*, 20(3), 1245–1253.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Proc. *Advances in Neural Information Processing Systems*, 30, 4768–4777.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. *Housing Economics and Public Policy*, 1, 67–89.

Matysiak, G. (2017). *Automated Valuation Models (AVMs): A brave new world?*. In Wroclaw Conference in Finance 2017, Wroclaw.

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.

Mohd, T., Jamil, S., & Masrom, S. (2020). Machine learning building price prediction with green building determinant. *International Journal of Artificial Intelligence*, 9(3), 379–386.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T. & French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, 21(4), 383–401.

Parker, D. (2016). *International Valuation Standards: A Guide to the Valuation of Real Property Assets*. Chichester: Wiley Blackwell.

Royal Institution of Chartered Surveyors (2015). *Our changing future: let's be ready*. RICS, London.

Royal Institution of Chartered Surveyors (2017). *Global Trends in Data Capture and Management in Real Estate and Construction*. RICS, London.

Schulz, R., Wersing, M., & Werwatz, A. (2014). Automated valuation modelling: a specification exercise. *Journal of Property Research*, *31*(2), 131–153.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, *13*(1), 1–44.