Multi-view Learning and Explainable Artificial Intelligence for Real Estate Appraisal

Der Fakultät für Wirtschaftswissenschaften der
Universität Paderborn
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
- Doctor rerum politicarum -
vorgelegte Dissertation
von

Jan-Peter Kucklick, MSc.
geboren am 02.04.1995 in Köln

2023

Abstract - English

Real estate appraisal, the house price estimation, is an essential but cumbersome task for selling, financing, or taxing a property. Information systems based on the hedonic pricing model arose, helping automate the process. These systems use classical house attributes like size and age in linear models. However, these systems fall short with limited predictive performance and cannot account for implicit factors like aesthetics embedded in unstructured data like images. Therefore, this work aims to improve real estate appraisal by leveraging multi-view learning models, which analyze multiple data types simultaneously. It has remained unclear which data sources to include and which modeling strategy to use to increase the predictive power. Furthermore, as most models are black boxes, explainability must also be inspected. Our experiments indicate that multi-view real estate appraisal models perform between 5.4% and 34% better than the baseline. A combination of geospatial data, exterior and streetview images was the best data combination, while multi-view neural networks were the most performant models. Explainable artificial intelligence helped to extract reasons for the predictions. In particular, we developed a new post-hoc method called Grad-Ram. From an economic perspective, comparing search and experience qualities revealed that both are essential for the price estimation. Lastly, a taxonomy describing the explainability of these systems was developed.

Keywords: Multi-view Learning, Explainable Artificial Intelligence, XAI, Real Estate Appraisal, Computer Vision, Deep Learning, House price

Abstrakt - Deutsch

Hauspreisevaluierungen sind ein essenzieller, aber arbeitsaufwändiger Bestandteil in Kauf-, Finanzierungs- oder Grundsteueraktivitäten. Informationssysteme basierend auf hedonischen Preismodellen helfen die Aufgabe zu automatisieren und fokussieren auf klassische Hauseigenschaften und lineare Modelle. Jedoch haben diese eine limitierte Vorhersagepräzision und können keine impliziten Preiseinflüsse aus unstrukturierten Daten wie Bildern bemessen. Das Ziel dieser Arbeit ist die Verbesserung der Hauspreisschätzung durch Multi-view Learning, welche verschiedene Datentypen gleichzeitig analysieren kann. Aktuell bleiben der Einfluss von Datenquellen und Modellierungsstrategie auf die Vorhersagegenauigkeit unklar. Zusätzlich muss die Erklärbarkeit der Algorithmen eruiert werden. Unsere Experimente zeigen, dass Hauspreisschätzungen basierend auf Multi-view Modellen zwischen 5.4% und 34% genauer sind. Eine Kombination aus geo-räumlichen Daten, sowie Außen- und Straßenbildern verbessert die Vorhersageperformance am stärksten, während Multi-view Neuronal Networks die performantesten Modelle sind. Erklärbare künstliche Intelligenz hilft bei der Identifizierung von Entscheidungsparametern, z.B. durch die entwickelte Grad-Ram Erklärbarkeitsmethode. Aus ökonomischer Sicht zeigt ein Vergleich von Such- und Erfahrungseigenschaften, dass beide essenziell für die Hauspreisschätzung sind. Schlussendlich wird eine Taxonomie für die Erklärbarkeit von Vorhersagesystemen konstruiert.

Stichworte: Multi-view Learning, Erklärbare künstliche Intelligenz, XAI, Hauspreisschätzung, Computer Vision, Deep Learning, Hauspreis

Contents

List of Figures

List of Tables

Acronyms

AOPC Area over the perturbation curve.

AVM Automated Value Models.

CAMA Computer Assisted Mass Appraisal.

CCA Canonical Correlation Analysis.

CNN Convolutional Neuronal Network.

CPU Central Processing Unit.

CV Computer Vision.

DL Deep Learning.

ETDP Extended Taxonomy Design Process.

GDPR General Data Protection Regulation.

GIS Geographical Information Systems.

GPU Graphical Processing Unit.

Grad-Cam Gradient-based Class Activations Maps.

Grad-Ram Gradient-based Regression Activation Maps.

HCML Human-Centered Machine Learning.

HCXAI Human-Centered Explainable Artificial Intelligence.

HIEF Holistic Interpretability and Explainability Framework.

HOG Histogram of Oriented Gradients.

HSV Hue, Saturation, Value.

ICE Individual Conditional Expectations.

IoT Internet of Things.

JQ3 Jourqual 3.

LIME Local Interpretable Model-Agnostic Explanations.

MAE Mean Absolute Error.

ML Machine Learning.

MLP Multi-Layer Perceptron.

MRI Magnetic Resonance Imaging.

MVC Multi-view Concatenation.

MVK Multi-view Kernel learning.

MVNN Multi-view Neuronal Network.

PDP Partial Dependence Plot.

POI Points-of-Interest.

RAM Regression Activation Maps.

RGB Red, Green, Blue.

RMSE Root Mean Squared Error.

SEC Search, Experience, Credence.

SIFT Scale Invariant Feature Transform.

SooS Spatial-out-of-Sample.

SWH Sliding Window Heatmap.

VHB Verband der Hochschullehrer der Betriebswirtschaft.

XAI Explainable Artificial Intelligence.

Part I.

Synopsis

1. Introduction

One basic need for humans is to find shelter - often provided by houses or condos. The value estimation of the place to live is essential for many tasks, including acquiring one. In this dissertation, we analyze and improve the computational models in the background. The subsequent chapters motivate the research objective and introduce the research questions answered.

1.1. Motivation

Real estate appraisal, the price estimation of a parcel, house, or apartment, is an essential task needed for many processes related to buying and selling, financing, or taxing the property (Law, Paige, & Russell, 2019; S. Peterson & Flanagan, 2009). While 12-16% of the Gross Domestic Product (GDP) of the U.S. is associated with processes around the house (B. Han, 2022; OECD, 2021), 3-5% are directly related to the sale process (National Association of Home Builders, 2020), indicating the economic importance of building, renovating, buying and selling houses. With the dominance of online platforms for a large variety of products, the search for a new home is evermore performed online. 97% of home buyers use online real estate platforms, like Zillow, the largest provider in the U.S., as an information source for a first search (National Association of Realtors, 2021).

Subject to any housing sale is a buying process. In general, the transaction process of property can be subject to information asymmetries, as the principal (buyer) has less information than the agent (seller), which can lead to market failure in the realm of the 'Markets for Lemons' (Akerlof, 1978; Ross, 1973). Especially for complex goods, which have information accessible prior to purchase, known as Search qualities, and information accessible only after the purchase, known as Experience qualities, the information asymmetries can reside (Nelson, 1970). One example of information asymmetries is the search on real estate online platforms like Zillow. Only Search qualities are depicted on the platform, but for example, the quality of the house or the safety of the neighborhood, characteristics that need to be experienced, are not available. This situation increases the advantage of the agent by having all this information. Making an accurate appraisal only with limited information makes the task even more challenging. In addition, appraisals are laborsome, as depending on the overarching process, they are performed manually by an expert (Koch et al., 2019; Sing, Yang, & Yu, 2021).

Intelligent systems named Computer Assisted Mass Appraisal (CAMA) (W. McCluskey, Deddis, Mannis, McBurney, & Borst, 1997) or Automated Value Models (AVM) (Sing et al., 2021) arose, aiming to reduce the overall workload in the appraisal process. Maybe the most prominent example of an AVM is the price estimate offered at the online real estate platform Zillow. It is often claimed, that price predictions are one of Zillows unique selling propositions (Poursaeed, Matera, & Belongie, 2018). Their price estimate, also called Zestimate®, is known for its precision because the prediction is within 5% of the sale price, 75% of the time (Zillow Group Inc., 2020).

Dominantly used for real estate appraisal is the hedonic pricing theory. The fundamental assumption in hedonic pricing is that the characteristics of a good contribute to the object's value. For houses, the number of rooms, age, or amenities like a fireplace are examples. Thus, the overall price is the sum of the value contribution of the characteristics (Lancaster, 1966; Rosen, 1974). A Linear Regression Model is often used to build the hedonic pricing model. Mathematically speaking, the hedonic pricing model and the Linear Regression have similar attributes of linearity and additivity. In addition, the Linear Regression is transparent because pricing effects (coefficients) can be measured, which can be used as an explanation for different user groups (Law et al., 2019). Recently, through the advances in Machine Learning (ML), more and more non-parametric models improved the predictive accuracy of the house price estimate (Kok, Koponen, & Martínez-Barbosa, 2017; Limsombunchai, 2004; Sing et al., 2021). This satisfies one essential criterion for using these systems in practice, as the predictive performance was reported of being important (Law et al., 2019; S. Peterson & Flanagan, 2009).

The existing literature focused on using variables typical for a house, like size, age, and amenities, which are called classical housing characteristics afterward (Limsombunchai, 2004; W. McCluskey et al., 1997; S. Peterson & Flanagan, 2009; Sing et al., 2021). Nevertheless, aspects that are also of importance in the pricing process of a house, e.g., style and aesthetics, are often omitted (Poursaeed et al., 2018; S. Zhang, Lee, Singh, & Srinivasan, 2022). However, this information may be included in unstructured data like images (e.g., exterior, interior, or satellite images). Because the information is not directly observable in the data, these variables are often called soft information (Liberti & Petersen, 2019). While they can not be directly analyzed, the advancements of ML, like Deep Learning (DL) algorithms, make it possible to use such implicit information (Poursaeed et al., 2018). Nevertheless, this leads to complex algorithms because images (three-dimensional tensor) and housing characteristics (vector) can not be easily fused. Consequently, current AVMs can not use these data sources simultaneously. The used plain vanilla[1] algorithms like Random Forest or Convolutional Neuronal Network (CNN) are specialized to analyze one data type only but can not combine multiple data types simultaneously.

1.2. Research objective and research question

This thesis, therefore, focuses on advancing existing real estate appraisal algorithms by incorporating such additional data sources. Therefore, a subfield in ML is used, which is called multi-view learning. It provides algorithmic solutions for leveraging multiple data sources (also called views) of different data types (Li, Yang, & Zhang, 2018; Zhao, Xie, Xu, & Sun, 2017). The research objective is:

Improving real estate appraisal by leveraging multi-view learning to incorporate multi-modal data.

Studies using this methodology in the context of real estate appraisal used different data sources and models (Bency, Rallapalli, Ganti, Srivatsa, & Manjunath, 2017; Bessinger & Jacobs, 2016; Law et al., 2019; Poursaeed et al., 2018). While a variety of data sources was used in the past, ranging from interior and exterior images to satellite images, no comparison of the predictive performance of different data sources exists. This led to the first research question:

RQ 1: What predictive power do different data sources (views) hold for real estate appraisal?

Furthermore, different modeling approaches were used to combine multiple data sources, ranging from a single to multiple models (Bency et al., 2017; Bessinger & Jacobs, 2016; Law et al., 2019; Liu et al., 2018). As pointed out, integrating various data sources in different formats is challenging, as most ML algorithms specialize in one data type (single view) only. One approach to deal with multiple views is to train one specific

[1] Plain vanilla is an English term to describe a basic flavor or usual configuration of an object. In computer science it refers to out-of-the-box (not customized) software.

algorithm per view and combine the predictions afterward. Other approaches extract features from the different views per ML algorithm, which are combined to be the input into another algorithm, predicting the overall target. Lastly, there are specialized DL algorithms, which create a latent feature space, used to predict the target variable. While these different methods compete, currently, no standard approach exists, and the influence of the model on the predictive power remains unclear. Thus, this raises the following question:

RQ 2: Which strategies in multi-view learning real estate appraisal lead to increased predictive performance?

Despite the requirement of giving precise estimates, AVMs must also be explainable to a user, either bound by law in decisions over humans (Kaminski, 2019) or because of user requirements in high-stakes decisions (Adadi & Berrada, 2018; Nussberger, Luo, Celis, & Crockett, 2022). For example, it might be necessary for a bank to explain why the loan for a house was rejected, while real estate agents using computer assistance to price houses might need to explain to their clients which characteristics influenced the house price (Law et al., 2019). While the hedonic linear regression possessed the qualities of being explainable and interpretable, through the use of ML, more and more black box algorithms were used. The ML's inner decision-making is obfuscated and thus complex to explain to humans because the algorithm provides no reason why the decision is made. To counteract this development, Explainable Artificial Intelligence (XAI) provided methods to explain the decision path of complex algorithms (Du, Liu, & Hu, 2019; Guidotti et al., 2018; Meske, Bunde, Schneider, & Gersch, 2020; Molnar, 2019). This leads to a third question:

RQ 3: To what extent can XAI increase the explainability in multi-view learning real estate appraisal algorithms?

To answer the research questions, two different research methods are used. First, we use computational experiments to create multi-view learning real estate appraisal models. These models relate after Gregor (2006) to Type 3 'Theories for prediction', as the predictive performance of strategies, data sources, and algorithms, as well as their explainability, is in the focus. Second, we create a framework, labeled after Gregor (2006) a Type 1 'Theory for analysis and description', to describe the AVMs' interpretability and explainability capacities.

Our results indicate that, on average, the multi-view learning models perform 11.7% better than the hedonic linear regression baseline. A combination of spatial data, exterior images and streetview images improved the predictive performance the strongest with an average of 15.4%, followed by satellite images with 14.5%. Overall, multi-view neural networks performed best with 14.3% improvement, followed by multi-view concatenation approaches with 11.5%. While concatenation approaches are the state-of-the-art, the suggested multi-view neural networks are on average 2.8 percentage points more precise, equaling approximately 8.000 USD per house. The used multi-view learning models are black boxes. However, some explainability can be gained by adjusting the modeling strategy and enhancing the models with XAI methods. For example, location and aesthetic aspects were learned from the models from image data. In addition, it was shown that these data types not only contain soft information but can also be used to transform Experience qualities to Search qualities. To be able to describe the kind of intelligent systems used, a taxonomy was built.

Our contribution is manifold. First, we extend the understanding of multi-view learning by testing how different data sources (views) can improve the predictive accuracy of AVMs in comparison to single-view AVMs only. Second, we compare different multi-view learning strategies for comprehending their contribution to the predictive power. By using multi-view neural networks, we challenge the widely used concatenation approaches, by (a) having a convenient single-stage model which provides predictions in an end-to-end fashion and (b) having better predictive performance. Third, we test the explainability of different multi-view learning approaches, advancing on strategies to improve it and provide a socio-technical framework for choosing an appropriate XAI method. In this light, we create a new XAI method for image regression problems to fill the methodological gap in related work. In addition, by providing a taxonomy, we enable the description of the interpretability and explainability capabilities in intelligent systems, especially of AVMs. Lastly, we inspect the role of Experience qualities in real estate appraisal and use geospatial analysis and DL to transform Experience qualities into Search qualities. This enables easier information acquisition and a shift in information asymmetries towards the principal, reducing the risk of a lemon market.

1.3. Thesis structure

Overall, six papers built the second part of the dissertation, enlisted in Table 1.1. These include three conference articles and three journal articles. One conference and one journal article are published as a single-author study. Five of the six papers published are ranked in the ‘Verband der Hochschullehrer der Betriebswirtschaft’ VHB Jourqual 3 (JQ3)[2]. The ranking of the articles included in this dissertation ranges from C to B.

In addition, a list of additional publications is summarized in Table 1.2. They include four conference outlets and one journal publication. These publications are pre-studies for the papers published about real estate appraisal. Furthermore, two relate to other topics, like data analysis in loyalty programs.

The dissertation is structured as follows: Part A comprises Chapters 1 to 6. Chapter 2 describes the foundations of real estate appraisal, deep learning for image data, multi-view learning, XAI and related work in multi-view real estate appraisal. Chapter 3 reflects on the research methods, while Chapter 4 summarizes the results. Chapter 5 discusses the results under various aspects of the research goal and concludes with implications. Chapter 6 ends the dissertation with limitations and an outlook. Each publication in Part B (Chapters 7 to 11) is one dedicated Chapter.

[2] The ranking can be found under: https://vhbonline.org/vhb4you/vhb-jourqual/vhb-jourqual-3

Table 1.1.: Publications included in this dissertation.

	Publication	**JQ3**	**Status**
1	**Kucklick, J.-P.**, & Müller, O. (2021). A comparison of multi-view learning strategies for satellite image-based real estate appraisal. In *The AAAI-21 Workshop on Knowledge Discovery from Unstructured Data in Financial Services.* Retrieved from https://aaai-kdf.github.io/kdf2021/assets/pdfs/KDF_21_paper_12.pdf		P
2	**Kucklick, J.-P.**, Müller, J., Beverungen, D., & Müller, O. (2021). Quantifying the impact of location data in real estate appraisal - a gis-based deep learning approach. In *Proceedings of the twenty-first European Conference on Information Systems (ECIS 2021), virtual, 14-16 June.* Retrieved from https://aisel.aisnet.org/ecis2021_rip/23/	B	P
3	**Kucklick, J.-P.** (2022b). Visual interpretability of image-based real estate appraisal. In *Proceedings of the 55th Hawaii International Conference on System Science (HICSS-55), virtual, 4-7 January* (pp. 1510-1519). doi: 10.24251/HICSS.2022.187	C	P
4	**Kucklick, J.-P.**, & Müller, O. (2023). Tackling the accuracy-interpretability trade-off: Interpretable deep learning models for satellite image-based real estate appraisal. *ACM Transactions on Management Information Systems (TMIS)*, 14(1). doi: 10.1145/3567430	B	P
5	**Kucklick, J.-P.**, Priefer, J., Beverungen, D., & Müller, O. (under review). Elucidating the predictive power of search and experience qualities for pricing of complex goods – A machine learning-based study on real estate appraisal. *Information Systems Frontiers.*	B	U
6	**Kucklick, J.-P.** (2023). Hief: a holistic interpretability and explainability framework. *Journal of Decision Systems*, 1-41. doi: 10.1080/12460125.2023.2207268	B	P

P: Published, U: Under review

Table 1.2.: Additional publications beyond the dissertation.

Publication	JQ3	Status
Kucklick, J.-P., & Müller, O. (2020). Location, location, location: Satellite image- based real-estate appraisal. *16th Symposium on Statistical Challenges in Electronic Commerce Research (SCECR).* Retrieved from https://arxiv.org/abs/2006.11406		P
Kucklick, J.-P., Kamm, M. R., Schneider, J., & Vom Brocke, J. (2020). Extending loyalty programs with BI functionalities. In *Proceedings of the 53th Hawaii International Conference on System Science (HICSS-53), Hawaii, 7-10 January* (pp. 168-177). Retrieved from http://hdl.handle.net/10125/63761	C	P
Kamm, M. R., **Kucklick, J.-P.**, Schneider, J., & vom Brocke, J. (2021). Data mining for small shops: Empowering brick-and-mortar stores through BI functionalities of a loyalty program1. *Information Systems Management*, 38(4), 270–286. doi: 10.1080/10580530.2020.1855486	C	P
Heuwinkel, T., **Kucklick, J.-P.**, & Müller, O. (2022). Using geolocated text to quantify location in real estate appraisal. In *Proceedings of the 55th Hawaii International Conference on System Science (HICSS-55), virtual, 4-7 January* (pp. 5744–5753). doi: 10.24251/HICSS.2022.700	C	P
Kucklick, J.-P. (2022a). Towards a model- and data-focused taxonomy of XAI systems. In *17th International Conference on Wirtschaftsinformatik (WI-22),virtual, 21-23 February* (pp. 1– 7). Retrieved from https://aisel.aisnet.org/wi2022/business_analytics/business_analytics/2	C	P

P: Published

2. Background

This chapter introduces basics for explainable multi-view real estate appraisal.

2.1. Real estate appraisal for tabular data

While real estate appraisal is a large field, first, the economic importance, user groups, and theories are introduced (Chapters 2.1.1, 2.1.2). Based on this summary, Chapter 2.1.3 describes technological developments in real estate appraisal concerning ML.

2.1.1. Real estate appraisal in the economic context

Real estate appraisal is the value estimation of houses, dwellings, or property land (Law et al., 2019; Limsombunchai, 2004). As mentioned earlier in Chapter 1.1, real estate is one vital industry for many countries in terms of economical impact. In contrast to trading other financial values like shares, real estate is characterized by a low trading frequency and no continuous cash flow (S. Peterson & Flanagan, 2009).

Appraisal is the basis for many tasks associated with real estate. Thus, different stakeholders are highly interested in it. First, real estate buyers and sellers rely on the price estimates as decision support in the buying or selling process, as well as real estate agents, who seek information to improve their sale pitch and negotiation (Bourassa, Cantoni, & Hoesli, 2010; Law et al., 2019; Limsombunchai, 2004). Banks and financial lenders use real estate appraisal as one criterion in their loan application process (Bourassa et al., 2010; Liu et al., 2018; Pagourtzi, Assimakopoulos, Hatzichristos, & French, 2003). Based on the evaluated value, the credit acceptance or rejection decision is made (S. Peterson & Flanagan, 2009). Municipalities and governments couple their property tax calculation on the appraised value of the house (W. McCluskey et al., 1997; Pagourtzi et al., 2003). Even not directly involved in the sales, payment, or taxation process of houses, economists and econometrics researchers are interested in appraisal as it has a dominant share in the economy and therefore is an essential part of their domain (Law et al., 2019).

Furthermore, estimating prices for houses as a decision aid is also popular in an online setting. Since online real estate platforms ease the property search, many buyers and sellers rely on the offered services. It is estimated that 97% of home buyers use the internet, like the online platform Zillow and its price estimate called Zestimate®, for a first search (National Association of Realtors, 2021; Zillow Group Inc., 2020). It is reported that the offered price prediction of Zillow is within 5% of the actual sale price 75% of the time (Poursaeed et al., 2018; Zillow Group Inc., 2020). Especially in online settings, where only the information available on the website is useable for searching and evaluating a product, information asymmetries can occur between the principal (buyer) and the agent (seller) (Ross, 1973). In the most extreme case, due to these information asymmetries, a 'market for lemons' is established, where good quality offers are not in demand anymore due to being indistinguishable from lower quality offers. The result is market failure (Akerlof, 1978).

Appraising a house is a complex task (Koch et al., 2019), as houses are non-standardized goods, and many aspects need to be considered in the value estimation process (Vanags, Geipele, Sarkans, & Usenieks, 2017). Therefore, experts often do this job in person. However, this is a time-consuming operation, especially when many houses need to be (re-)evaluated (Koch et al., 2019). In addition, it is important that experts keep their price estimation consistent between houses (Sing et al., 2021).

Using a decision support system based on quantitative data, statistics, and ML might help decrease the work effort per appraisal and standardizes the estimation process (Sing et al., 2021). Such systems are named Computer Assisted Mass Appraisal (CAMA) (W. McCluskey et al., 1997; Peña, Fuentes, Cervera, & Hernández, 2012) or Automated Value Models (AVM) (W. J. McCluskey, Daud, & Kamarudin, 2014; W. J. McCluskey, McCord, Davis, Haran, & McIlhatton, 2013; Sing et al., 2021). The prediction based on the quantitative model can either be used to automate the appraisal process fully or to support a decision maker in the decision process by providing an objectively produced anchor (Koch et al., 2019; Peña et al., 2012; S. Peterson & Flanagan, 2009; Sing et al., 2021). Looking at usage statistics from appraisers revealed that while 97% of the appraisers questioned conducted at least once an appraisal in-person, only 7% have tried AVMs. Moreover, 45% of all appraisers stated that they are very uncomfortable with AVMs, being the least popular choice of appraisal methods (National Association of Realtors, 2022).

2.1.2. Hedonic real estate appraisal

In general, there are many approaches to predict house prices and sales, including methods focused on individual houses and price index-based time-series models[1] (Baldominos et al., 2018; Hill & Scholz, 2018; Limsombunchai, 2004). Focusing on the evaluation of individual house prices, the housing characteristics like size and age as independent variables, and the current value of a house as a dependent variable matter. Real estate appraisal, therefore, is grounded in the Hedonic Pricing Model. This theory was proposed in the 1970s by Lancaster (1966) and later on refined by Rosen (1974) in the 1980s. The underlying assumption is that an object's value is determined by its characteristics and usefulness. Thus, the utilities should be separated and estimated individually, primarily by gathering real-world data and measuring the value differences by variation in the descriptive characteristics of the object. The total value of an object is the sum of the value contribution of the constitutional characteristics, expressed as:

$$y = \sum_{i}^{I} \beta_i \cdot x_i \tag{2.1}$$

where y is the total value of the object (e.g., a house), x_i is a utility (e.g., number of rooms), β_i is the value contribution of x_i, summed over all characteristics I. As Equation 2.1 is, in fact, in a linear model, often a Linear Regression is used to estimate the hedonic pricing model. The strength of this theory is that the value contributions of characteristics are objectively measured, creating a very good domain understanding, similar to the decision process of stakeholders (Pagourtzi et al., 2003). Furthermore, as predominately a Linear Regression is used, the algorithm itself is very interpretable with the statistical measurement of the coefficients β.

Besides the hedonic pricing theory, other methods exist to appraise a home. One method that uses the house's characteristics to determine the price is the Comparable Pricing Approach. When performed by hand, the appraiser selects similar houses to the house to appraise, determines their similarity, and makes the final estimate by weighting the prices according to the similarity (Pagourtzi et al., 2003). This procedure could also be automatized through ML by using the k-nearest-neighbors algorithms, which

[1] Time-series models use an index to predict based on a time-based structure the mean house prices (for a specific type of house, e.g. family home), for the upcoming 6, 12, 18 month (Baldominos et al., 2018; Hill & Scholz, 2018). In this thesis, we solely focus on appraisal, the evaluation of individual houses, and therefore disregard research about house price indexes.

computes the similarity based on the distance of characteristics. In contrast to the methods focused on the house's characteristics, several approaches are related to finance or geospatial analysis. Financial approaches are investment or profit based. Both have a cash-flow-centric view of the property. In the investment-based evaluation, rent measures the real estate value, as the rental value is the monthly or yearly equivalent usage of a property paid to the landlord (owner). The rent can be seen as the return of investment for buying the property. In case the property does not relate to a house but an industrial company or service firm, the property is seen as a production factor (e.g., the rooms in a hotel), and the potential cash flow in terms of profit is calculated by using the turn-over and deducting any costs. Property can also be seen from a developmental standpoint, where the usage of the land and dwelling determines the value. The value is determined by geographical factors. Different use cases are compared, with the goal of finding the most suitable for the land at hand. Thus, the starting point is to imagine the land would be unused, and a use has to be determined. Consequently, it is implicitly assumed that the current usage of the parcel is not the best and can be improved. Often geographical factors (distance to POIs, lot dimensions, physical attributes) instead of housing characteristics are considered. This method can help to create more attractive spaces and thus might be attractive for municipalities. Lastly, there are methods related to the cost of replacement. In terms that the property type is rarely traded or sold and is unique or very special in terms of usage, so that there are no market observations to determine a value by hedonic or comparison approaches, the value of the property is the costs that would be needed to rebuilt or replace this current property (Pagourtzi et al., 2003). For this work, we mainly focus on the hedonic pricing approach, as this is primarily used for the type of property that is inspected (family homes) as well as because of its unique properties to measure the value contribution of the characteristics of a house that can be used to explain the prediction made (Law et al., 2019; Limsombunchai, 2004).

In contrast to pricing methods, another economic theory must also be considered related to product search and evaluation. As introduced at the beginning of the chapter, the sales process, especially in an online setting, has to face information asymmetries when analyzed from a principal-agent setting (Ross, 1973). Despite the side-effect of possibly failing markets (Akerlof, 1978), theories explaining the cost-effort and information-acquiring process in a buying process emerged. In particular, this is the Search, Experience, Credence (SEC) theory. While initially proposed by Nelson (1970) focusing on the differences between Search and Experience qualities, the theory was extended by Darby and Karni (1973) to Credence goods. First, the theory categorized the whole good as a Search or Experience good, while later on, the theory was applied to individual characteristics of goods (Klein, 1998). In general, the SEC theory describes the properties of a good, depending on the point of time, when the information is available, and which costs are associated with them (Figure 2.1). The overall costs are the product costs plus the search costs for the product (Klein, 1998). Search goods have a majority of qualities where information can be accessed prior to purchase (Nelson, 1970). So the information is relatively easy to acquire, resulting in low search costs. Acquiring information by purchase (experience) is in this case costlier than the search for information. Classic examples are run-of-the-shelf consumer goods like socks or furniture (Iacobucci, 1992). In contrast, experience qualities are qualities that need to be experienced so that these can be evaluated. Acquiring information about these properties before the purchase is costlier in terms of effort and time. Because of the expensive search for information, these products would usually be evaluated by sampling - by repeatedly buying and experiencing them. If the experienced good does not meet the requirements (e.g., taste), another brand or 'manufacturer' is tried out (sampled from all available products). Examples are washing powder, the quality of a service (rental car), or the taste of food (Iacobucci, 1992). Sometimes the hurdles for sampling are additionally lowered by offering free trials so that customers are engaged in experiencing the product (Klein, 1998). Credence goods have qualities that can not be accessed because a comparison is impossible. For example, a consultation with a lawyer or a medical operation falls under the credence category. One can not fight the same lawsuit twice using different lawyers or remove the same kidney from one's body by different doctors, so an insufficient amount of data can be observed to make comparisons between the lawyers or doctors (Darby & Karni, 1973; Iacobucci, 1992).

In this context, goods with having Search and Experience qualities are interesting, because both information acquisition ways need to be fulfilled. Real estate is one of these products, according to the survey of Iacobucci (1992), where a condo is described with Search as well as Experience qualities. Setting the SEC theory back in the market and principal-agent-context, information asymmetries exist not only because of the agent's superior knowledge but also because of the increased search costs for the principal for the Experience qualities like the safety of a neighborhood or the condition of a house. In particular, while consumer goods like washing powder or food can be evaluated by sampling, due to the high acquisition costs of real estate and a complex purchase process in some countries, experiencing the qualities through a purchase is often impossible. This setting might also be amplified on real estate platforms because only the information displayed can be used to evaluate the product, but touching, smelling, or experiencing the house is not possible. Despite this hurdle, information technology can be an enabler to simplify product evaluation according to Klein (1998). Different technologies transform experience criteria to search quality. We posit in Paper 5 (Kucklick et al., under review), that through newly emerging technologies like ML and Geographical Information Systems (GIS), Experience qualities can be transformed to Search qualities by making them measurable and expressive from newly available data sources (unstructured data). This helps to capture aspects like safety of the neighborhood, aesthetics or noise levels, which needed to be experienced before to be included in the decision and pricing process. As now these information can be made expressive (hard information), it can be used in AVMs to increase their predictive performance. The connection between the principal-agent theory, SEC theory, and the work of Klein is visualized in Figure 2.2

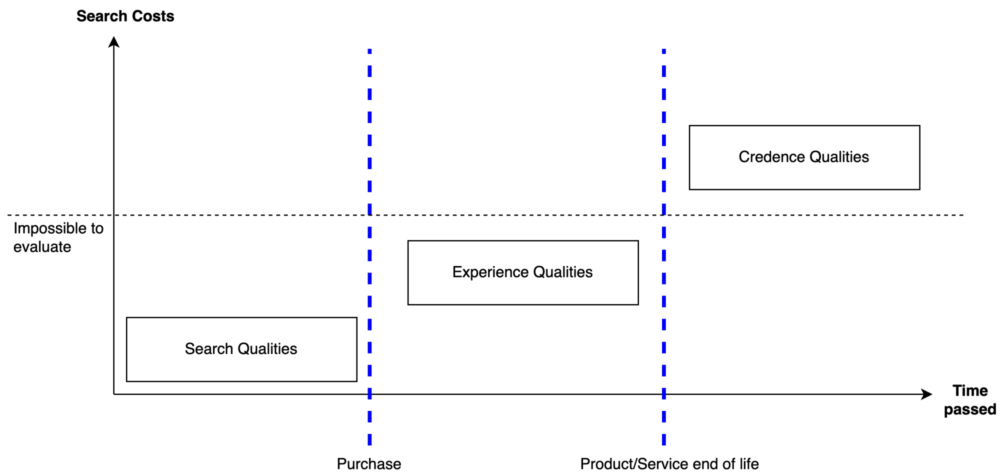


Figure 2.1.: Theoretical visualization of SEC Theory. Search qualities are available before the purchase at relatively low costs. To collect Experience qualities, the product or service has to be purchased plus consumed to get information about this characteristics. Credence qualities are nearly impossible to evaluate due to extremely high costs to acquire.

While the SEC theory and linked microeconomic observations are other lenses to view real estate appraisal, this line of theory is not in focus in appraisal, arguably because it only describes the search process. However, it does not directly help to make pricing decisions. Nonetheless, as we will see in the following chapters, the SEC theory will help to uncover interesting insights in real estate appraisal.

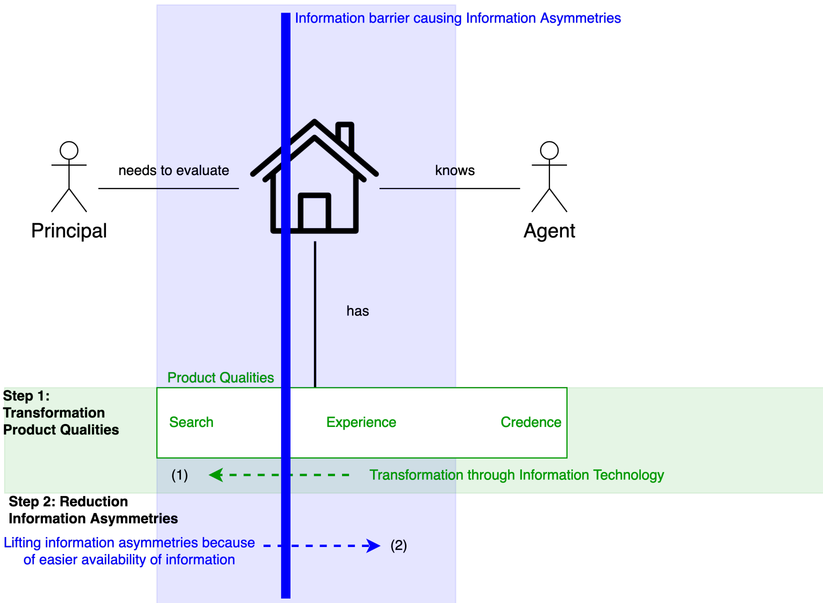


Figure 2.2.: Principal-Agent Theory and SEC Theory in the context of real estate appraisal (Darby & Karni, 1973; Klein, 1998; Nelson, 1970). First, Experience qualities can be transformed through technology to Search qualities (Green line and box) by mining newly available data sources. These information are available for the principal. Second, this helps to reduce information asymmetries (blue box) to move the information barrier away from the principal (bold blue line). Icon Kiranshastry (n.d.) by Flaticon.com.

2.1.3. Machine Learning for real estate appraisal

Previously, economic motivations for real estate appraisal have been discussed. Subsequential, we will start analyzing technical ML aspects in the subsequent chapters.

Due to the similarities of the hedonic pricing model and the linear regression, this analysis technique was often used in the past (Law et al., 2019; Limsombunchai, 2004). Variables describing the utilities are often fact-based and are represented in a vector or tabular format, including, for example, the size (in square feet, number of (bath)rooms), age, special amenities like fireplaces or pools, and the neighborhood the house is located (Kok et al., 2017; Limsombunchai, 2004; W. McCluskey et al., 1997; Potrawa & Tetereva, 2022; Sing et al., 2021). To improve such algorithms, under the realm of ML, more and more advanced models are used to raise the predictive performance (accuracy of the value estimates) by models like Decision Trees, Random Forest, Gradient Boosting, or dense neural networks (Kok et al., 2017; Limsombunchai, 2004; W. J. McCluskey et al., 2014, 2013; B. Park & Bae, 2015; Potrawa & Tetereva, 2022). Many of these models are non-parametric; thus, no assumption about the model and the variable interdependency needs to be made. Consequently, the models are, in mathematical terms, very flexible to fit a given feature space by automatically modeling non-linearities and interaction effects (James, Witten, Hastie, & Tibshirani, 2013; Liao & Varshney, 2021). This suits the necessary criterion in real estate appraisal of having exact estimates (Pagourtzi et al., 2003; S. Peterson & Flanagan, 2009).

Up to this point, the described real estate models focused only on using tabular data, which hold the classical real estate variables like size, age, and amenities. The following section will introduce Computer Vision (CV) and image analysis techniques. In the following chapters, we will see how to use this information for real estate appraisal.

2.2. Deep Learning for image data

Convolutional Neuronal Network (CNN) are one type of Deep Learning (DL) algorithm, often used to analyze images (Kraus, Feuerriegel, & Oztekin, 2020; LeCun, Bengio, & Hinton, 2015). DL excelled in the last years because the algorithms are representation learners, so they can handle the extraction of various features from raw and unstructured data. This functionality sets them apart from classical ML algorithms, which can only learn on extracted and preprocessed variables, where a part of the information gets lost in the transformation. For example, while CNNs use the 'raw' image, CV with ML requires to extract variables, e.g., the color distribution by transforming the image into Hue, Saturation, Value (HSV) representation (Joblove & Greenberg, 1978), or Histogram of Oriented Gradients (HOG)[2] features, which are a summary of the image in a vector format (Dalal & Triggs, 2005). For example, the green color distribution extracted from satellite images was used in real estate appraisal as a proxy for urban density and vegetation (Kostic & Jevremovic, 2020). Nevertheless, learning from the raw features like pixels can increase the predictive performance of an algorithm because more information is kept, making CNNs a popular choice for CV (Kraus et al., 2020).

2.2.1. The nature of image data

Before explaining CNNs in-depth, more details on raw image data should be explained. An image is a three-dimensional tensor of image height, width, and the number of

[2] HOG creates a histogram of the direction (orientation) of found edges in (subspaces of) the image. HOG can be used to transform the image tensor into a vector representation. HOG features, for example, are used to train a classical ML model (Support Vector Machine) to detect pedestrians in images (Dalal & Triggs, 2005).

channels (Figure 2.3). The value for each cell is typically between 0 and 255. The number of channels is three for a colored image, representing the amount of Red, Green, Blue (RGB) colors at each pixel. In Figure 2.3, the area in the red bounding box in the original image is visualized as the RGB matrix, which the computer uses. Nevertheless, different kinds of images have different numbers of channels. For example, black-and-white images have one channel. Magnetic Resonance Imaging (MRI) produces images with many channels, each representing one scan at one depth of an object. In other images, like satellite images, there are additional channels to the RGB, capturing laser or radar wavelengths to understand more aspects of planets, like atmospheres, amount of water on the surface or vegetation (L. Chen et al., 2018; Donaldson & Storeygard, 2016; LeCun et al., 2015).

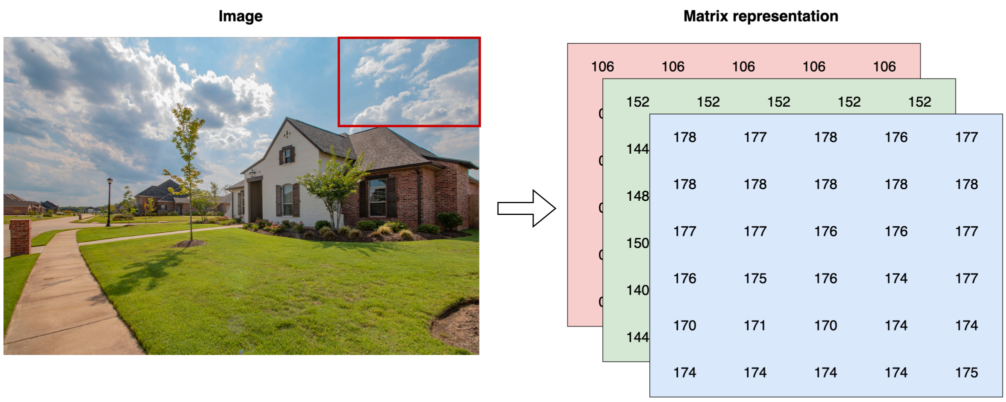


Figure 2.3.: Visualization of a three-channel image (RGB) as image and matrix. The area in the read bounding box in the original image is transformed to the RGB matrix on the right. Original image provided by George (2020).

Images differ from structured data as the order of the variables in structured data does not matter for the outcome. Assuming that a house can be described by three variables (number of rooms, size, and neighborhood), in Figure 2.4 (A), two houses are described. When the order of the variables is changed, the table on the right still describes the identical houses as the table on the left. Thus, an algorithm trained on the left and another on the right table should have the same results when it is deterministic. For example, in a linear regression, only the coefficient order will be changed according to the change in the order of the variables.

However, this property changes in images. An image has a lot of spatial dependencies. One pixel or even a small local region is connected to other regions. When the order of regions changes, as in Figure 2.4 (B), the image and its meaning change. In the middle visualization of Figure 2.4 (B), it can be seen that one can not detect the window anymore, for the case when the pixels at the middle height of the image are randomly shuffled along the width. When a random order of 'columns' (pixels along the image's width) is applied to the full height, as visualized in the right image, the house, front yard, and pavement can no longer be identified.

Part A: Tabular data

Tabular data with different column orderings

Original Table

Rooms	Size	Neighborhood
2	75	A
3	85	B

=

Random column order

Size	Neighborhood	Rooms
75	A	2
85	B	3

Part B: Image data

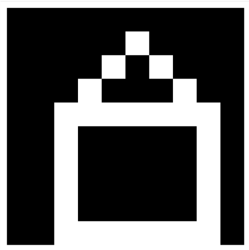
Spatial dependency in images

Figure 2.4.: Differences between structured and unstructured (image) data. Part A indicates structured data and Part B shows image data. The left image is the original image depicting a house. In the middle visualization, the pixel order is randomly shuffled along the image width for a narrow patch in the image height. Right visualization depicts the original image, when a random order of columns along the image width is applied. It can be seen that for this image, the object (house) is not identifiable anymore. Original image provided by George (2020).

Classical ML algorithms, like Decision Trees, Random Forests, Support Vector Machines, or DL algorithms like multi-layer perceptrons, require data to come in a vector format. These algorithms cannot preserve image properties. The only way to use image data in these algorithms without preprocessing would be to flatten the image so that for each pixel and each channel, the values are concatenated in a long vector. In such a format, the algorithm no longer knows which pixels belong together. To preserve the property that the pixel order matters and that pixels near each other are correlated and belong to the same 'object', e.g., a blue pixel with blue pixels nearby may represent the sky in the image (Figure 2.4), a convolutional filter is learned from the raw data (pixels) (LeCun et al., 2015). In CV, a convolutional operation filters the image for

Image (to be analyzed)

Visualization

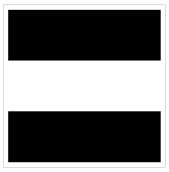


Mathematical representation

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	255	0	0	0	0
0	0	0	0	255	0	255	0	0	0
0	0	0	255	0	0	0	255	0	0
0	0	255	255	255	255	255	255	255	0
0	0	255	0	0	0	0	0	255	0
0	0	255	0	0	0	0	0	255	0
0	0	255	0	0	0	0	0	255	0
0	0	255	0	0	0	0	0	255	0
0	0	255	255	255	255	255	255	255	0

Kernel (Pattern that is filtered)

Visualization



Mathematical representation

0	0	0
1	1	1
0	0	0

Figure 2.5.: Example image and kernel. For demonstration purposes, the image shows a simplified house.

specific patterns (Kraus et al., 2020).

2.2.2. Building blocks of Convolutional Neural Networks

In 'traditional' CV methods, the filter is hand-crafted. The filter consists of a matrix (kernel) containing the pattern to filter. Therefore, the weights of the kernel (elements of the matrix) have specific values. For example, values of 0 and 1 detect if the pixel is inactive or active for this pattern. The kernel is sled over the image, and for each movement, the dot product between the kernel and the part of the image where the kernel 'looks', is calculated (Goodfellow, Bengio, & Courville, 2016). The output of this transformation is often called a feature map. For example, the simplified house in Figure 2.5 should be filtered for horizontal patterns. In this example, a three-by-three kernel is applied to the image, where the output shows how 'horizontal' the patterns are in the image. The procedure is visualized in Figure 2.6 including the resulting feature map. For the first two movements of the kernel, one can see that the output is zero for each movement, as in the receptive field of the kernel (fields highlighted in red), no horizontal patterns are depicted. The kernel size, the stride size (how many pixels the kernel moves each time), and if the kernel can also be moved over the edge of the original image (padding) are all hyperparameters that can be set and designed. For example, the filter for a horizontal pattern is one part of manual edge detection, as in Canny (1986).

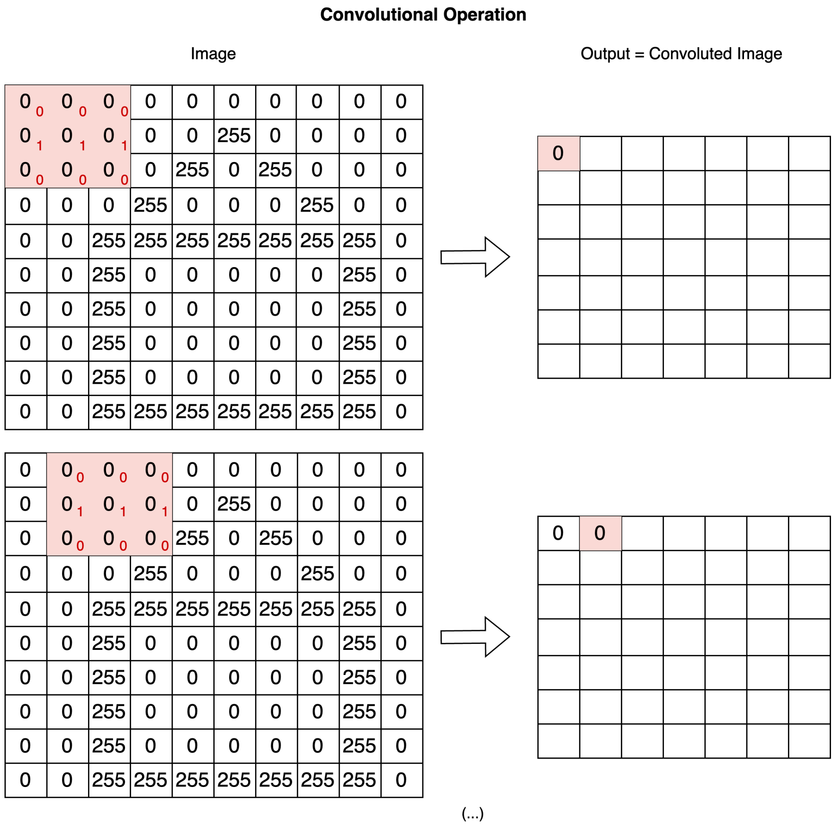


Figure 2.6.: Convolutional operation on image with kernel presented in Figure 2.5. The receptive field is highlighted through red cells in the matrix. Red numbers indicate the kernel weights.

Applying a convolution to a multi-channel input follows the same logic described above, except that there are multiple kernels, one for every input channel. These multiple kernels comprise a convolutional filter together. The feature map is just the sum of the outputs of all kernels. As each kernel is independent of the others in the filter, they can also have different weights. However, the parameters like size, stride size, and padding stay the same.

Many convolutional filters create a convolutional layer. In CNNs, the filter's weights are learned by backpropagation in contrast to designing them per hand, as in the example above. CNNs often stack convolutional layers together so that the feature map of one layer is the multi-dimension input representation for the next layer. In contrast to classical neuronal networks like Multi-Layer Perceptrons (MLPs), where all the neurons are connected to all neurons in the next layer, building dense connectivity, only a small patch in the one layer is connected to a small patch in the next layer. This creates sparse connectivity. This allows (a) to learn very fine-granular local patterns in the data and (b) does require much fewer weights, reducing the computational complexity of the whole network (Goodfellow et al., 2016; LeCun et al., 2015). Moreover, CNNs also bear an additional property. The convolutions perform a so-called weight sharing. When a filter detects a certain pattern, it will apply the same weights (included in the kernel) to the pattern anywhere in the image (Goodfellow et al., 2016; LeCun et al., 2015). This is a desired property because local patterns, e.g., a window can be at multiple locations in a house picture (e.g., left or right, or on the second floor), should consistently be rated with the same influence. In MLP, the weights are learned individually for each input. For a flattened image with a pixel being x_1 and another pixel being x_n, which could belong to the same object (e.g., window), get different weights, w_1 and w_n, where w_1 does not equal w_n.

Despite sparse connectivity and weight sharing, two additional properties make CNNs suitable for analyzing image data: Translation invariant and hierarchical learning. When, for example, the door design slightly changes (e.g., the doorknob is mounted higher or mounted to the middle), the CNN should still be able to filter this pattern as a door. Being robust to minimal pattern changes is created using pooling layers (Goodfellow et al., 2016). A pooling layer replaces the output of a convolution by calculating a kind summary statistic for nearby values of the feature map (Figure 2.7). The summary statistic often relates to the mean or maximum value. Similar parameters such as the pooling size, the stride size, padding, and the method for the summary statistics can be set.

Lastly, hierarchical learning uses the depth of the neuronal network because many CNNs are comprised of tenths or hundreds of layers (He, Zhang, Ren, & Sun, 2016a). As the input features for one layer are the extracted patterns of the previous layer, neuronal networks learn patterns in a hierarchical way. The early layers of a CNN often detect simple patterns such as color compositions and edges, which can be combined into textures, which form subparts, and small objects, which form the object itself (LeCun et al., 2015). For example, a CNN-detecting house might first learn a composition of many vertical, horizontal, and diagonal patterns, which are combined into textures like the structure of the wall. Features at that level, e.g., roof form, wall patterns, windows, and doors, are then combined to learn the pattern of a house. Learning hierarchically and the composition of detecting local patterns (convolutional layers) and summarized

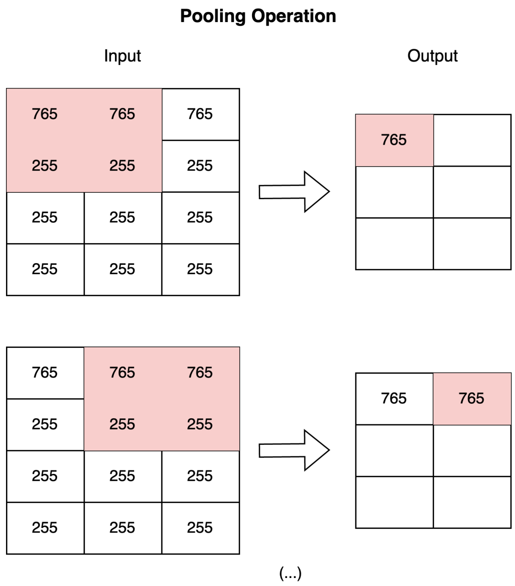


Figure 2.7.: Visualization of a pooling operation. In this example, max pooling (using the maximum value) is performed. Input is the output of a convolution. Output is the pooled image

patterns (pooling layers) is very similar to the visual processes in the human body. The visual cortex also consists of cells, which can extract local patterns (simple cells), and complex cells, capable of summarizing the patterns of the simple cells. To provide computers the ability to understand visual data was inspired by the processes in the human brain (Goodfellow et al., 2016).

Finally, a CNN consists of many stacked convolutional layers, which may be enhanced by using pooling layers between the convolutional layers. In addition, some network architectures use a MLP at the top (layers towards the output) to weight and combine the detected patterns and to provide a vector-based output (see Figure 2.8). The earliest CNN trained with backpropagation was designed by LeCun et al. 1989 and classified the numbers in hand-written zip codes from letters to make them machine-readable and sortable. CNNs increased in importance strongly after 2010 because the training of the networks, especially with large image resolution, is time and resource-complex. However, training speeds could be improved by using Graphical Processing Units (GPUs) as hardware accelerators instead of relying on Central Processing Units (CPUs). Speaking in a simplified manner, GPUs make use of heavy parallel computation with their hundreds or thousands of cores (Chollet, 2017). In addition, many improvements concerning the neural network architecture have been made over the more than 25 years CNNs have existed. For example, methods were introduced to train deeper networks (more layers) (He et al., 2016a), stacking of convolutions was applied to reduce the number of parameters, and thus the computational complexity of the algorithms (Simonyan & Zisserman, 2014), or new structures of layers within CNNs were introduced to automatically find the best convolution sizes (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016).

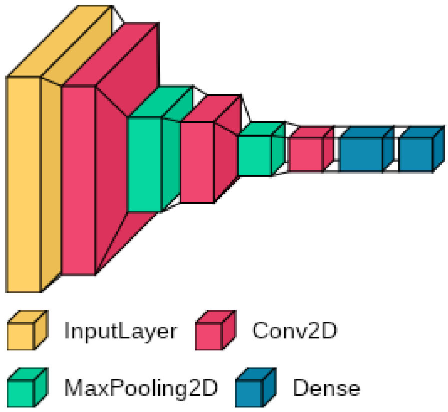


Figure 2.8.: Architecture of Convolutional Neural Network. Created with Visualkeras (Gavrikov, 2020)

With the building blocks of CNNs, real estate images, e.g., exterior images or satellite images of the location, could be used to directly regress the price when the output layer is adjusted of the CNN. This would resemble a single-view approach, as only one type of data input would be used. Interestingly, CNNs, being able to analyze data with spatial relationships very well, e.g., images, can not be logically applied to tabular data. As pointed out in Figure 2.4, the order of the columns in tabular data are not of importance. However, the CNN assumes a local structure (ordering) in the input data. Thus applying a CNN to tabular data leaves an ill-defined algorithm (Kraus et al., 2020).

While Chapter 2.1.3 focused on applying ML models to tabular real estate data, this chapter introduced CV and image data. However, the question arises, what needs to be done when multiple data types arise simultaneously, and which are all necessary to analyze a problem? The following section about multi-view learning introduces concepts to solve this challenge.

2.3. Multi-view learning

In the previous chapters, we have seen how real estate appraisal is performed on tabular data with the hedonic pricing model. While the algorithms were enhanced by ML, they typically focused only on classical house attributes. Meanwhile, CNNs emerged that are specialized in analyzing image data. Nevertheless, convolutions always assume a spatial relationship, which does not exist in tabular data. Thus, to combine tabular data and image data within one algorithm, other strategies must be applied. The newly emerged field of multi-view learning provides strategies for combining data with different data types for a model. The following chapter gives an introduction to this field.

2.3.1. An introduction to multi-view learning

The digital transformation of companies, business models, and processes not only led to opportunities to create value and to improve services and products but also created a high data availability (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013; H. Chen, Chiang, & Storey, 2012; Vial, 2019). Enforced by the Big Data paradigm, data is stored from multiple sources in large amounts (Volume), with various formats (Variety), and is created at high speed (Velocity), as often the data-generating process is linked to a data stream (Fosso Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015; McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). With the Internet of Things (IoT) and mobile devices, mini-computers and sensors are everywhere and connected to generate, process, and react based on data. Furthermore, with social media and a growing number of business models driven by services rather than physical products, data has undoubtedly become one important resource in the 21st century (Bharadwaj et al., 2013; Davenport & Patil, 2012; Fosso Wamba et al., 2015; McAfee et al., 2012). Information needs to be connected to support complex decision processes by combining all the different data sources (H. Chen et al., 2012; Zhao et al., 2017). However, this turns out to be challenging because ML models, especially those developed for structured data, expect the data to come in the first normal form. Therefore, each observation is a row, and each variable is a column, where each value needs to be atomic to be machine readable (James et al., 2013). Nevertheless, this behavior contradicts the Big Data paradigm, as a variety of data formats need to be combined, where information can have tabular characteristics (vector), but also unstructured data like an image is linked to it, which is represented as a 3D tensor (height, width, channels). However, classical ML algorithms can not use inherently unstructured data types without heavy pre-processing. In this case, a lot of signal is lost. Therefore, multi-view learning is a field where modeling strategies are researched to combine information from various sources (also called views), where each view can also be based on another data type (e.g., tabular, textual, sequential, visual, audio, geospatial) (Li et al., 2018; S. Sun, 2013; C. Xu, Tao, & Xu, 2013; Zhao et al., 2017). Interestingly, while the views have distinct feature sets, these feature sets can be generated by the same data type, e.g., color vs. texture of images (C. Xu et al., 2013) or multiple images from different perspectives (Yan, Hu, Mao, Ye, & Yu, 2021). Figure 2.9, gives an example of a single vs. multi-view set-up, where tabular data or images alone are used for the single view. In the multi-view example, tabular and two types of image data are combined.

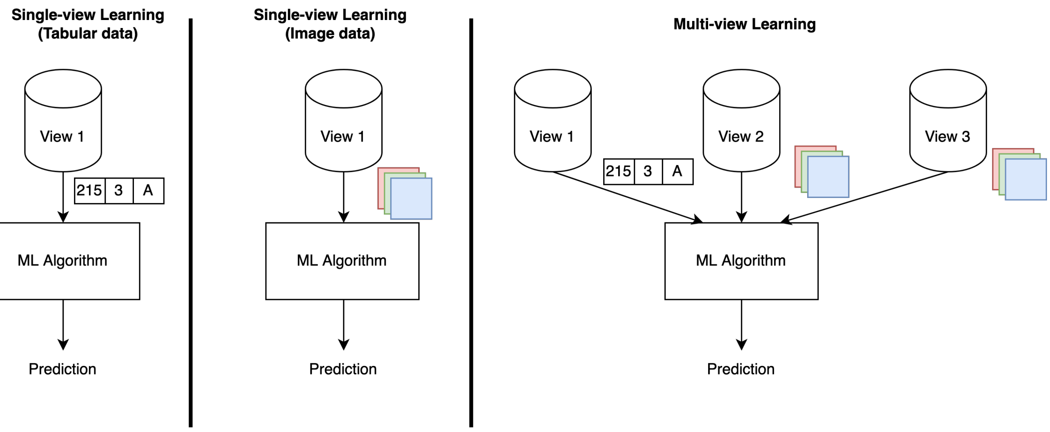


Figure 2.9.: Single vs. multi-view application. The single-view model uses just one data type (either tabular or image data), while the multi-view model uses tabular data in combination with two images.

Multi-view learning can be separated into two sub-fields: multi-view representation fusing and multi-view representation alignment. Multi-view representation alignment aims to create a joint feature space representing the knowledge derived from the different views. Therefore, each view is preprocessed by a specific mapping function, and the differences between the views in the created feature space are minimized (Li et al., 2018; S. Sun, 2013; C. Xu et al., 2013). The created joint feature space can be seen as a new representation of knowledge. In contrast, multi-view representation fusing aims to create a new feature space based on the theory that new views contain additional information that has not been captured beforehand. This additional information should improve the predictive accuracy of an ML algorithm (Li et al., 2018; C. Xu et al., 2013). The generated feature space is extended by new features of other views. An example technique for multi-view representation alignment is Canonical Correlation Analysis (CCA) and various distance-based optimization processes (Li et al., 2018; S. Sun, 2013).

For multi-view representation fusion, multi-kernel learning, multi-view concatenation, and multi-view neural networks are exemplary analysis techniques (Figure 2.10) (Li et al., 2018; C. Xu et al., 2013; Yan et al., 2021). In brief, for a Multi-view Kernel learning (MVK) strategy, individual models are trained for each view (data input). The predictions are combined, e.g., by majority vote or average weighting, yielding the overall prediction (Figure 2.10, (A)). In the case of using Multi-view Concatenation (MVC), individual models (Model 2 and Model 3) are trained to extract features from the views so that the variables can be concatenated or joined. Often the resulting data are structured tabular data. With these variables, a prediction model is trained, making the overall prediction (Figure 2.10, (B)). Lastly, in a Multi-view Neuronal Network (MVNN), one model (a neural network) is created, that can handle multiple inputs due to the multi-branch architecture, consisting of four subnetworks. In this example, a MLP and two CNNs are forming the three branches. The information is fused together, and another MLP learns from the fused features to make the overall prediction. All branches are trained simultaneously, and the created model with its four subnetworks predicts the target variable (Figure 2.10, (C)). A detailed overview of multi-view fusion techniques can be found in Paper 1 (Kucklick & Müller, 2021).

Applications of multi-view learning are widespread in real-world applications because data is often collected from various sources and measuring methods since a single data source can not holistically capture all information (S. Sun, 2013; Zhao et al., 2017). Multi-view learning is applied in multi-media analysis, medical classification, video surveillance, social network analysis, machine translation, and recommendation systems (Li et al., 2018; S. Sun, 2013; Yan et al., 2021).

2.3.2. Delimitation between multi-view learning and multi-modality

Recently, another concept called multi-modality arose in competition with multi-view learning (Baltrušaitis, Ahuja, & Morency, 2019), which was also recently discussed in real estate appraisal (Azizi & Rudnytskyi, 2022). First, multi-modality will be briefly defined, and then the similarities and differences to multi-view learning will be described in the next paragraph.

Multi-modality is motivated by the senses of humans. Frequently, humans combine information from multiple senses, e.g., visual impressions and sounds, to make a judgment. Thus, multi-modality is the use of distinct data sets and algorithms capable of analyzing this information to better represent the phenomena or object (Baltrušaitis

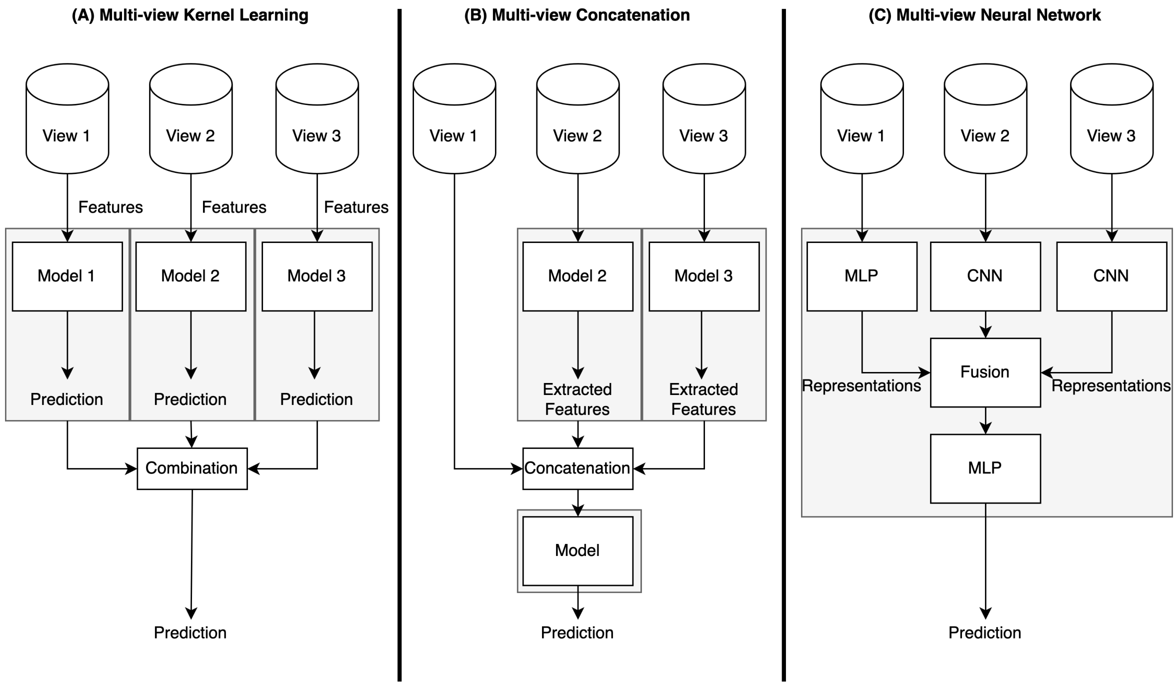


Figure 2.10.: Different multi-view learning strategies. Multi-view kernel learning (Subplot A), multi-view concatenation (Subplot B), and multi-view neural network (Subplot C). Grey boxes indicate individually trained models. While multi-view kernel learning and multi-view concatenation use three models in this example, the multi-view neural network is one model consisting of four subnetworks.

et al., 2019). Based on this definition, the features must contain different data types, e.g., audio and video. In the taxonomy paper of Baltrušaitis et al. (2019), five different subcategories are identified in multi-modality: representation, translation, alignment, fusion, and co-learning. These match roughly the multi-view learning categories, where translation, alignment, and co-learning are linked to multi-view presentation learning, fusion is associated with multi-view representation fusion, and representation, being the overarching characteristic, is linked to both. Typical multi-modality applications are audio-visual speech recognition, visual question answering, multi-media event detection, emotion detection, multi-media retrieval, and generation, to name a few (Baltrušaitis et al., 2019; Castellano, Kessous, & Caridakis, 2008).

Similarities and differences between multi-view learning and multi-modality are summarized in Table 2.1. While both concepts use multiple data and feature sets, they differ in their motivation and data types selection: Multi-view learning is rooted in the idea of getting as many possible representations of an object or situation as possible. Therefore,

Table 2.1.: Comparison of Multi-view learning and Multi-modality

	Multi-view learning	**Multi-modality**
Source	Multiple feature sets	Multiple feature sets
Origin & Motivation	Extend representation of object by using ubiquitous data sources	Mimic human perception (Sensory differences)
Data types	Can vary	Must vary
Subcategories	Alignment Fusion	Representation, Translation, Alignment, Co-learning Fusion

ubiquitous data sources are acquired and used. In contrast to this technical motivation, multi-modality is inspired by nature. Humans get a better perception through using different senses (e.g., combining sound and smell). In this case, sense equals in the ML world different data types. Consequently, an ML model can be improved if different nature of data are used. The differentiation in the origin and motivation of the two theories also lead to differences in the possible data types. While multi-view learning models can have different data types but do not have to (e.g., multiple images in the work of Yan et al. (2021)), the multi-modal applications must have different data types. The differences in the subcategories are based on different, more fine-granular definitions, however, they seem content-wise identically. Based on the described similarities and differences, it seems that although they are different concepts, they are strongly related. Multi-view learning having fewer data types restrictions seems to be the more generic term, as every multi-modal application is a multi-view application, but not every multi-view learning application is also multi-modal.

2.4. Explainable Artificial Intelligence

This chapter focuses on the technical details (Chapter 2.4.1) and social aspects (Chapter 2.4.2) of Explainable Artificial Intelligence (XAI) to provide a holistic social-technical perspective. ML models have greatly advanced in predictive performance, and many technological innovations in the field of CNNs and multi-view learning emerged. Nonetheless, one desired criterion for an algorithm is to be understandable for a user. Mostly this need originates from the requirement to follow algorithmic decisions (Martens & Provost, 2014). Users do not want to follow blindly any system which they can not understand and which can not reason why particular decisions were made (Adadi &

Berrada, 2018). Nussberger et al. (2022) link psychological theories. Humans can understand, learn and predict the behavior of an ML model through the use of XAI. This gives the user the feeling of being in control, which in term establishes trust.

2.4.1. Technical aspects of Explainable Artificial Intelligence

In general, XAI refers to a research field and methods that help explain algorithms (Barredo Arrieta et al., 2020; Meske et al., 2020). The next paragraphs introduce different XAI dimensions, which are summarized in Figure 2.11.

Type of Understandability	Interpretability	Explainability	
Type of ML Model	Inherently Interpretable	Black-box	
Type of Explainability	Local	Global	
Type of Explainability Method	Model Inspection	Outcome Explanation	Model Explanation
Type of Evaluation	Metrical	Experiment	

Figure 2.11.: Technical dimensions of Explainable Artificial Intelligence.

Type of understandability

In previous research, interpretability and explainability were used as synonyms when it was referred to the understandability of algorithms. However, interpretability and explainability, although related, have different focuses. Explainability defines the power to extract a reason of a model, why a particular prediction was made, e.g., Which factors lead to an increase in the real estate price? (Meske et al., 2020). Thus, with this focus, only variables and the direction and sometimes strength of their effect on the target variable are of interest (Barredo Arrieta et al., 2020). Contrasting, interpretability shifts the focus from the reasons to the decision process of the algorithm. It is equal to asking the question, how does the algorithm make decisions? Consequently, an algorithm is perceived as interpretable when humans can understand the mathematical optimization process and algorithm formula (Barredo Arrieta et al., 2020). Specifically, there are different notions of interpretability. Lipton (2018) defines that interpretability is built on three concepts, namely algorithmic transparency, decomposability, and simulatability, where the latter requires the former to be established. Algorithmic transparency defines how comprehensive the model is in mathematical terms. Complex optimization processes, like mini-batch-based backpropagation within high dimensional non-linear loss functions, result in very complex loss landscapes that often can not be understood by a human, especially when compared to a single linear regression, where the optimization process can be visualized. Furthermore, some algorithms are not deterministic, making it difficult to compare the results and to see if an optimal solution in the training process is reached (Lipton, 2018). Decomposability describes the property of the algorithm so that inputs (variables), weights, and calculations can be separated and individually understood. Simulatability refers to the ability to contemplate the algorithm's behavior by the user. The human brain has limited capacities, and some algorithms are too complex (by having too many parameters) so that their decision process can not be 'simulated' in the human brain (Barredo Arrieta et al., 2020; Lipton, 2018). For a deduction of the influence of interpretability on explainability, see Paper 6 (Kucklick, 2023), where the connection is made explicit.

Type of ML Model

In general, there are two ways to provide an explainable algorithm. The first one is to use inherently transparent (also called: interpretable) algorithms that the user can understand and analyze. The second one is the use of black box algorithms in combination with post-hoc explainability methods. Black box algorithms are named this way because their internal decision-making process is obfuscated so that a user can neither interpret the decision path of the algorithm nor explain why the algorithm made the prediction (Du et al., 2019). Examples of black box algorithms are Random Forests, Support Vector Machines, and Neuronal Networks, including CNNs (Asatiani et al., 2021; Barredo Arrieta et al., 2020; Lipton, 2018).

Type of Explainability

Furthermore, for both approaches, explainability can be provided on a global or local scale (Du et al., 2019; Molnar, 2019). Local explainability focuses on single observations and provides factors for the decision of this particular instance. A question that could be answered with local explanation is: which part of the house is increasing the value? In Figure 2.12, each black line visualizes a house and how the price changes by changes in size. In contrast, global explainability tries to analyze the behavior of the whole model, e.g., how do house prices, in general, change when the size increases? The global behavior is visualized as the red line in Figure 2.12.

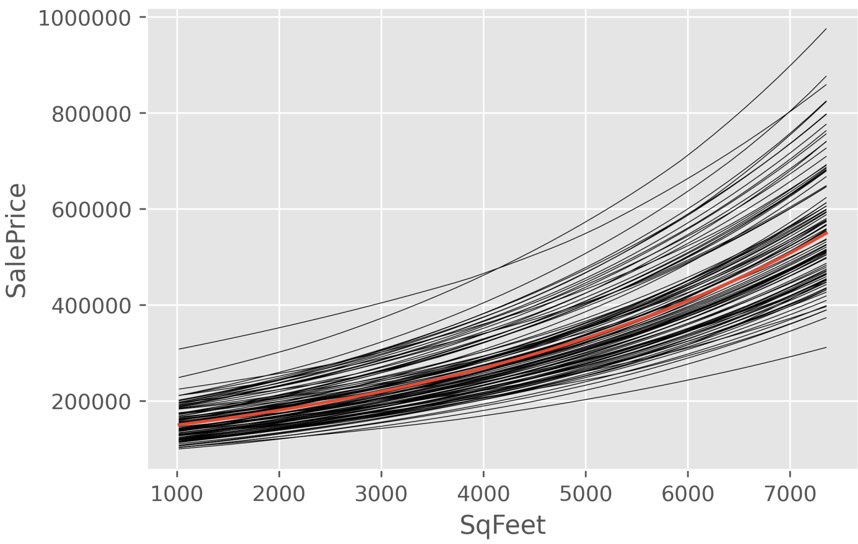


Figure 2.12.: Partial Dependence Plot - Individual Conditional Expectation visualization on data from Kucklick and Müller (2023). Y-Axis visualizes the sale price as being dependent on the size of a house (SqFeet) on the X-Axis. Partical Dependence Plot (PDP) visualized in red, and individual houses are highlighted with the black Individual Conditional Expectation (ICE) lines.

Accuracy-Interpretability Trade-off

When there is the option of using already (inherently) transparent models in contrast to black box models and XAI methods, the question arises, why are not inherently transparent models the standard? It is because there is a trade-off between interpretability and accuracy (Gunning et al., 2019). More interpretable algorithms like a Linear Regression or a Decision Tree are perceived to perform worse than black box models. Thus, performance needs to be sacrificed for explainability. While some authors find a linear ordering between the understandability of algorithms and the performance (Asatiani et al., 2021), others identify either a non-linear ordering (Herm, Heinrich, Wanner, & Janiesch, 2023; Wanner, Herm, Heinrich, & Janiesch, 2021) or no trade-off at all (Gosiewska, Kozak, & Biecek, 2021; Rudin, 2019). Especially for ML problems with structured (tabular) data, Rudin (2019) found that often inherently interpretable algorithms with clever variable selection are on par or even outperforming black box algorithms. In the context of why black box algorithms are predominately used, Liao and Varshney (2021) state these algorithms often do not need effortful and time-consuming feature engineering.

Type of Explainability Method

In light of XAI, different post-hoc explainability methods have been designed. Guidotti et al. (2018) summarize these methods and give a good overview, separating three general types of post-hoc explainability methods. These are model inspection, model explanation and outcome explanation. Methods described as model inspection are used to explain the relationship between one or two independent variables and the dependent variable. These methods provide a similar interpretation to one to two coefficients of a Linear Regression. Examples of this method are Partial Dependence Plot (PDP) or Individual Conditional Expectations (ICE) (Friedman, 2001; Goldstein, Kapelner, Bleich, & Pitkin, 2015; Molnar, 2019). In contrast to model inspection methods, model explanation methods try to explain the whole algorithm at once. One example technique is extracting decision rules, e.g., using RxREN (Augasta & Kathirvalavakumar, 2012). In the case of outcome explanations, an explainable model that can mimic the behavior of the black box model is used. Famous examples are Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro, Singh, & Guestrin, 2016) or Gradient-based Class Activations Maps (Grad-Cam) (Selvaraju et al., 2017).

In addition, there are more dimensions to consider in XAI. Explanations also depend on the data type used in the model (Barredo Arrieta et al., 2020). For example, to make sense of images, a graphical explanation is required, highlighting areas that contributed to the decision (Figure 2.13), while, for example, to find the strength and direction of a variable, a (regression) table seems to be a more suitable choice. For some methods, especially those where the model's behavior based on one or two factors is simulated (model inspection), a graphical explanation might be suitable too. PDP or ICE are visualized in figures, where the x-axis demonstrates the change of the independent variable, and the y-axis denotes the outcome (Figure 2.12). The cognitive fit theory can be linked to selecting the displaying method (Vessey & Galletta, 1991). If a method is perceived as helpful (fast information requirement, low error rate) for a task, it is determined by the match with the problem. For so-called spatial tasks, seeing patterns and trends, graphs are more suitable than tables, while tables are the display choice when reading a single value is of importance.

Explainability for images

For describing image explanation methods in more detail (Figure 2.13), the example and corresponding CNN from Chapter 2.2 are used. The CNN is used as an input into an XAI method, creating a weight matrix, which can be visualized as a heatmap. In this example, dark blue areas in the heatmap indicate that this part of the image does not affect the prediction, while light blue areas indicate a small and green and yellow areas indicate a medium effect. Red areas highlight a strong effect. In our simplified house example, one can see that the model relied only on triangular and horizontal lines to detect the house but missed the vertical lines (walls). This example indicates the possibility of monitoring the model's logic and, in case of systematic error, correcting it. The heatmap is often also overlayed (merged) with the original image to directly indicate the important parts, which makes an interpretation easier[3]. Different XAI methods have different procedures how to create the weights in the heatmap (Molnar, 2019), with methods having a strong focus on classification (Montavon, Binder, Lapuschkin, Samek, & Müller, 2019; Selvaraju et al., 2017; Simonyan & Zisserman, 2014). However, XAI methods for explaining continuous outcomes for regression problems (e.g., house prices) are sparse, as only two methods are available, namely sliding-window-heatmap and Regression Activation Maps (Z. Wang & Yang, 2018; Yang, Chen, & Chou, 2018).

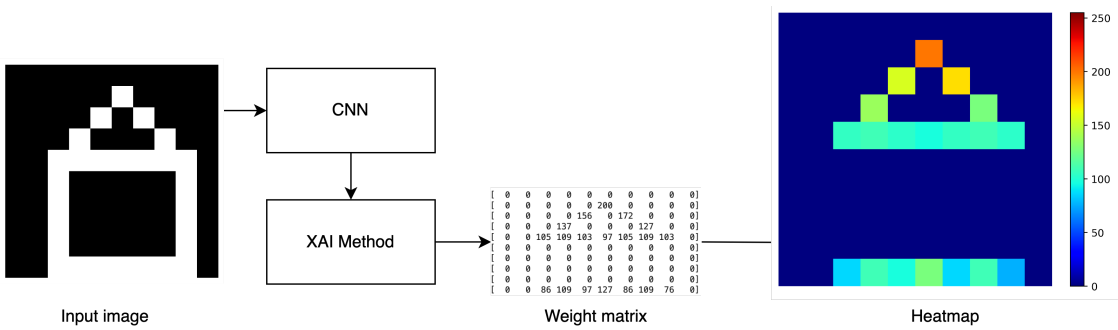


Figure 2.13.: Schematic visualization of an explanation of image data. The CNN is used by an XAI method to create a heatmap. Dark blue colors indicate no influence on the prediction. In this example, green and yellow colors indicate a medium influence and red colors visualize a strong influence. XAI methods differ in how the heatmap is created.

Limitations of XAI methods & Type of evaluation

While XAI methods can help the user better understand the factors of a decision (Wastensteiner, Weiss, Haag, & Hopf, 2021), post-hoc explainability methods come with disadvantages. As these methods often are models of the algorithm to explain, one has to guarantee that these are faithful. Therefore, it has to be ensured that the explainability method is sensitive to the model and the data, as well as a check is necessary if the identified reasons of the model match their estimated effect (Adebayo et

[3] In this example, as the example image contains only black and white pixels, but no gray areas, an overlay is not possible.

al., 2018; Alvarez-Melis & Jaakkola, 2018; Samek, Binder, Montavon, Lapuschkin, & Müller, 2016; Tomsett, Harborne, Chakraborty, Gurram, & Preece, 2020). To test these properties, either metrics can be calculated like the Faithfulness Score or the Area over the perturbation curve (AOPC), or user experiments are performed in which the overlap of a human-made explanation and the post-hoc explanation is measured (Alvarez-Melis & Jaakkola, 2018; Ribeiro et al., 2016; Samek et al., 2016; Selvaraju et al., 2017).

2.4.2. Socio aspects of Explainable Artificial Intelligence

Despite the technical aspects of XAI described above, social dimensions must also be included in the design and selection of XAI methods. As the explainability method is often used to improve the decision process of a human, based on the algorithmic prediction and its reasoning, explainability is highly task, domain, and user-specific (Adadi & Berrada, 2018; Liao & Varshney, 2021; Martens & Provost, 2014; Mohseni, Zarei, & Ragan, 2021). A summary of social aspects of XAI discussed in the following subchapters is provided in Figure 2.14.

Usage Aims	Justification	Controling	Knowledge generation	Debugging & Improvement
User Group	End User	Manager	Data Scientist	

Figure 2.14.: Social dimensions of Explainable Artificial Intelligence.

Usage Aims of XAI

In general, there are four motivations for using XAI. These are to justify the decision made by the algorithm, to be in control of the algorithm by understanding wrongful and flawed reasoning, to improve the algorithm by a better understanding, or to generate new knowledge (Adadi & Berrada, 2018). Precisely when the decision made by the algorithm has a potentially strong influence, so-called high-stakes decision, like having costly, fatal or unethical consequences like in finance, medicine or law, explainability is absolutely required (Adadi & Berrada, 2018; Du et al., 2019; Guidotti et al., 2018; Law et al., 2019; Rudin, 2019). Other studies indicate that not only the type of stakes in the decision influences the users' desire to have explainable algorithms (Nussberger et al., 2022). In fact, in situations where the ML algorithms make decisions about access to scarce resources and act like gatekeepers, understandability is highly valued too. Presumably, the user (human) wants to check if the allocation is performed fairly

(Nussberger et al., 2022). Interestingly, the wish for XAI is not influenced by the predictive accuracy of the algorithm. Thus, despite having an accuracy of 60% or 90%, high-stakes decision or gatekeeping functions of ML release the desire of having understandable algorithms for users. Nevertheless, when the user was confronted with a trade-off situation between accuracy and interpretability, accuracy was preferred (Nussberger et al., 2022). This contradicts the before-stated desire for explainability without an explaining theory. Nonetheless, it portrays the danger of an 'answers first, explanation second' attitude because, in this case, severe consequences for users produced by algorithms can occur before having the ability to understand and fix the issue (Nussberger et al., 2022). In this context, also administratives worldwide saw the need to describe the necessity for explainable ML algorithms in legislative form. One example is the General Data Protection Regulation (GDPR) from the European Union (Kaminski, 2019; Samek & Müller, 2019). In particular, ML models need to be explainable when the decision has a significant impact on the data's subject, which was regulated in the 'right to explanation' (Recital 71).

User groups of XAI

Thus, an ML system has different user groups, ranging from the end user via a manager to the developer of an ML system. Considering these users, they have different backgrounds and requests on an algorithm (Martens & Provost, 2014; Mohseni et al., 2021). While the end-user might need to justify her decision, a data scientist might require insights to improve the algorithm. Consequently, other techniques of these user groups will be used, e.g., local vs. global explainability or outcome explanations vs. model inspection. A shift of the requirements for a data scientist role is described in Belle and Papantonis (2021), where multiple iterations are performed on the XAI methods until all the requirements are satisfied.

7-Gaps model

The 7-gaps model by Martens and Provost (2014), visualized in Figure 2.15 provides a theoretical approach to consider the algorithm, its performance, and different user groups in relation to each other. The overall assumption of the 7-gap model is that a user trusts a model and adapts the decision (acceptance) when they understand the model's reasoning and when the argumentation is aligned with their knowledge (user's mental model) and the real world's effects. Consequently, the overall aim is to close the gaps between these concepts. Thus, explanations have two functions: (a) improving the decision system (ML Model) and (b) improving the knowledge of the users. The gaps in the model account for a bidirectional influence. Gaps one to four refer to improving the systems, while gaps five to seven refer to knowledge generation. Gap one describes the difference between the decision system (ML Model) and reality. When gap one is large, the algorithm is inaccurate and not representing the real world. By improving the model performance, gap one can be closed. Gaps two to four are the gaps between a user's mental model and the algorithm, split by user role. Different user groups require different kinds of explanations. For example, a data scientist needs to understand why false prediction of the system occurs to debug and improve the algorithm, while the manager is responsible for fair decisions and needs to understand the ML model's overall behavior (Martens & Provost, 2014). An explanation can help to trust the model because the user's mental model and the algorithms are more aligned. However, a better understanding of the model can also create a feedback loop and improve the ML model based on the user's knowledge. Gaps five to seven are between the users' mental models and reality. The explanation provided by the algorithm can improve the user's understanding of reality. In other words, explainability methods can provide previously unknown insights (knowledge generation) that help the user better understand reality.

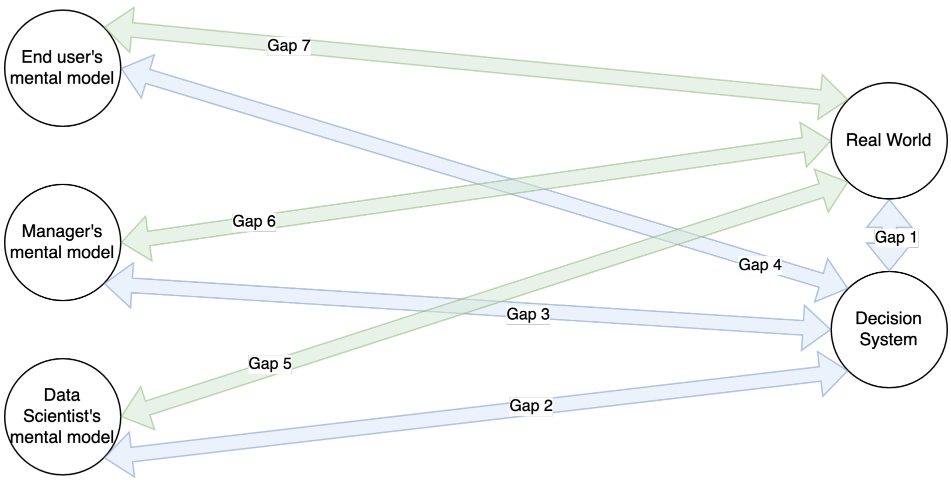


Figure 2.15.: 7-Gaps framework according to Martens and Provost (2014). Blue arrows resemble the relationship between user groups and the decision model. Green arrows highlight the relationship between the users' mental models and reality based on the explanation. The overall aim is to eliminate all gaps in the system.

Human-Centered Explainable Artificial Intelligence

Criticism arose on XAI methods from a socio-technical view because these methods are often designed by experts (data scientists) but are not understandable for other user groups except these experts (Liao & Varshney, 2021). To counteract this development, a new field called Human-Centered Explainable Artificial Intelligence (HCXAI) was created (Förster, Klier, Kluge, & Sigler, 2020; Mohseni et al., 2021). In particular, the authors in this literature stream suggest development processes strongly aligned to the user's requirements and capabilities, including multiple design rounds, so that the end-user is integrated from the early steps. In a wider understanding, Human-Centered Machine Learning (HCML) was recently defined as methods for creating, understanding, improving, and criticizing ML-based systems from perspectives that balance technological innovation as well as human, social, ethical, and political concerns (Chancellor, 2023). In the author's understanding, the HCML is thus the superordinate concept of HCXAI, which bears many similarities to considering technology from a socio-technical perspective in information systems after Bostrom and Heinen (1977).

While a wide range of applications uses XAI to support users in their decision-making (Adadi & Berrada, 2018; Guidotti et al., 2018), real estate appraisal as a financial application also benefits from these methods. Especially as this application makes high-stakes decisions with a large financial value and many different stakeholders rely on appraisal models, requirements can be complex. While XAI methods were already used for house price predictions (e.g. De Nadai & Lepri, 2018; Law et al., 2019) before this work, there is a call for providing XAI methods for real estate image data (Law et al., 2019). Therefore, the next chapter will introduce the use case of multi-view real estate appraisal in more detail.

2.5. Related work in multi-view real estate appraisal

First, in Chapter 2.5.1, we argue why multi-view real estate appraisal is necessary, while in Chapter 2.5.2, we compare the existing research in detail and link the articles to our research objective.

2.5.1. Motivation for multi-view learning in real-estate appraisal

Usually, facts about a house can easily be represented in a format ML can leverage because variables can be expressed as numerical or categorical values. Nonetheless, a house's price does not only depend on these easy-to-measure variables. 89% of real estate buyers state that photos are a very helpful information source, which is a greater percentage than the house characteristics (86%) (National Association of Realtors, 2021). For example, house prices are also influenced by information about the appearance of a house and the yard (Elam & Stigarll, 2012; Johnson, Tidwell, & Villupuram, 2020), the perceived luxury level (Poursaeed et al., 2018) and greenery in the area (Donovan & Butry, 2011; Law et al., 2019). However, these variables are hard to express in a structured format. In finance, the former variables would be described as containing hard information, while the latter variables are described as soft information (Liberti & Petersen, 2019). Soft information, in contrast to hard information, is difficult to express in numerical or categorical variables, as the information is often available between the lines (in unstructured data), e.g., in a transcript of a clients financial assessment, instead of directly accessible numbers, e.g., savings amount or interest rates. However, soft information, despite being harder accessible, bears unused signals, potentially improving the accuracy of a ML model.

Using multiple data sources with different data formats seems necessary to leverage the information described above, leading to a multi-view learning application. The soft information about the house's appearance, style, and coziness or the yard's state is often represented in images, which should be combined with the facts (housing characteristics) to determine the price. The importance of using multiple data sources (multi-view learning) is highlighted in two recent publications in real estate appraisal (Naser, Serte, & Al-Turjman, 2021; Wei et al., 2022). Both publications stress that ML and image analysis advances allow necessary aesthetic details about the house to be incorporated into the pricing. Furthermore, Wei et al. (2022) predict, because of the increasing availability of data by Big Data and open data policies, that the acquisition and analysis of these data sources will be of rising importance. In contrast to Naser et al. (2021), who focus on the value increasing by visual data (images), Wei et al. state that also spatial data like Points-of-Interest (POI) from Geographical Information Systems (GIS), or mobility patterns of inhabitants derived from IoT sensors, GPS, and smart cards, will add additional signals to the analysis. Consequently, the data focus will shift from housing characteristics like size, age, and location towards data (e.g., GIS, images, POIs) collected with other purposes like modeling traffic flows, navigation, or spatial analysis of neighborhoods to improve the currently used appraisal models.

Nevertheless, while using more available data (variables) and more sophisticated (ML) algorithms could improve the predictive performance of AVMs, factors about explainability should not be neglected. As multi-view learning described in Chapter 2.3 can lead to complex modeling procedures, explainability will be limited. Nonetheless, this conflicts with real estate appraisal being a financial application. Making such high-stakes decisions (Guidotti et al., 2018) and due to being used in loan application processes being part of a ML-gatekeeper process to scarce (financial) resources (Nussberger et al., 2022), users reported that explainability is extremely important. Consequently, this could lead to an accuracy-interpretability trade-off (Gunning et al., 2019), where a weight between predictive power and understandability needs to be found. In particular, in the light of the 7-Gaps model (Martens & Provost, 2014) and HCXAI (Mohseni et al., 2021), users of intelligent systems have a desire to understand the intelligent system, building trust in the algorithm's capabilities and thus increase the usage, while also having the chance to expand the existing knowledge. Consequently, explainability is a very important factor of an intelligent system too. In particular, as 7% of appraisers have used AVMs and do not feel comfortable with the usage (National Association of Realtors, 2022), combining ML with XAI has the potential to increase the acceptance of AVMs in the real estate community.

2.5.2. Comparison of related work in multi-view real estate appraisal

Despite the theoretical views on real estate appraisal and the already deducted important topics about multi-view learning and explainability, in the next paragraphs, the state of research about image-based real estate appraisal is summarized. We start with articles until 2018, which is the starting point of this thesis, to motivate our research (Table 2.2) and continue with the ongoing development and trends in this domain. In the following tables, papers using images in combination with housing attributes are compared based on their used data type, the multi-view learning strategy, and the usage of XAI. The multi-view learning strategy notation is based on the work of Kucklick and Müller (2021), who compared and listed different modeling strategies in the real estate domain.

By reviewing the data views included in image-based real estate appraisal, several studies use various information from GIS variables, exterior images, street view data[4], interior images and satellite images in combination with the classical housing attributes (Ahmed & Moustafa, 2016; Bency et al., 2017; De Nadai & Lepri, 2018; Law et al., 2019; Poursaeed et al., 2018; You, Pang, Cao, & Luo, 2017; Zhao, Liu, Kuang, Chen, & Yang, 2018). While data sources are manifold, the broad research in this field raises the question of what value different views have in predictive performance and which ones should be included in a multi-view real estate appraisal model. Significantly, the

[4] In contrast to exterior images, which show the exterior of the house (sometimes even from the curb), street view image face in the driving direction and thus cover mainly the road and neighboring houses from the viewpoint (Law et al., 2019).

scarce research about satellite images is a promising research field — this forms and supports RQ1.

Table 2.2.: Paper-Concept-Matrix of related work in image-based real estate appraisal until 2018.
Abbreviations: Multi-view kernel learning (MVK), Multi-view concatenation (MVC), Multi-view neural networks (MVNN). Visual explanation: Regression denoted by R, Classification by C. X marks if the feature or category applies to the paper. Brackets indicate a limited application of the category.

Paper	**Data**								**Model**	**XAI**	
	House attributes	**GIS attributes**	**Satellite Image**	**Exterior Image**	**Streetview**	**Interior Image**	**Map**	**Building plan**		**Tabular**	**Visual**
Elam and Stigarll (2012)	X			X					MVC*	X	
Ahmed and Moustafa (2016)	X			X		X			MVC		
Bessinger and Jacobs (2016)	X			X					MVC	(X)	
Bency et al. (2017)	X		X						MVC		
You et al. (2017)				X					MVNN		
Poursaeed et al. (2018)	X			X		X			MVC		
Zhao et al. (2018)	X	X			X				MVC	X	
De Nadai and Lepri (2018)	X	X			X				MVC	X	
Liu et al. (2018)	X			X					MVNN		(R)
Law et al. (2019)**	X	X	X		X				MVNN	X	
Summary (Total: 10)	9	3	2	6	3	2	0	0	MVC (7) & MVNN (3)	5	1

(R): In the paper, important parts in the image are highlighted by a bounding box. With the description on the paper, it remains unclear if an XAI method was applied or if the bounding box was created by other means (e.g., weights of the neural network).
**: The paper of Law et al. (2019) was available as a pre-print in 2018.

In addition to the multiple views, also various multi-view learning strategies exist. Elam and Stigarll (2012) do extract the soft information in the form of aesthetic scores form of a house curb appeal score (How beautiful is the house seen from the street?) and a landscape appeal score (How beautiful is the landscape (greenery, trees, garden, etc.)

around the house?). The extraction is based on a human rating, and no computer-driven image feature extraction is involved. This results in a small sample size of 75 houses, as this process is work-intensive. Law et al. (2019) argue this issue, as feature extraction is labor intensive, especially when specific variables need to be extracted. Therefore, a data set needs to be created in the first step to train an ML algorithm to automate this task. For example, in the work of Poursaeed et al. (2018), ten thousand images were labeled into different luxury classes so that an algorithm could extract this information. By combining this information with house attributes, a Multi-view Concatenation (MVC) strategy is used. Ahmed and Moustafa (2016), used a similar approach, where interesting information from images using classical computer vision methods (Scale Invariant Feature Transform (SIFT)) was extracted. SIFT summarize the image tensor in a vector that can be joined with tabular data. However, one must admit that SIFT is not specialized in extracting specific real estate patterns, as hard-coded filters are used. Other authors trained so-called Multi-view Neuronal Network (MVNN), a special form of a neural network (Law et al., 2019; Liu et al., 2018; You et al., 2017). This network has multiple input branches, and each branch is specialized for an input data type. The learned features are fused later on in the same model to predict the price. The whole algorithm is trained end-to-end so that no additional models and, therefore, no data labeling are necessary. This can be reached because all views are directly regressed on the price. This method has the advantage that real estate-specific patterns are learned during training by adapting convolutions by backpropagation. However, Bency et al. (2017) report that these algorithms are difficult to optimize. Maybe this is one reason only three of ten papers used this method. The large variety in the modeling strategy, from human-based feature extraction, via domain-specific knowledge extraction (e.g., luxury levels) to technical complex multi-input neural networks, raises the question of which modeling strategy leads to more predictive performance supporting RQ2.

One valued characteristic of the hedonic pricing model and the used linear regression model is its interpretability. This utility is often used in research and practice because the marginal willingness to pay (coefficients of the linear regression) contains valuable information (Law et al., 2019). However, with more sophisticated ML algorithms, models have become black boxes (James et al., 2013), leading to a trade-off between the model's accuracy and its interpretability (Gunning et al., 2019). Interpretability is not only necessary for the end-user but also for managers and data scientists because it helps to extract new knowledge, monitor and (quality) control the algorithm and raise the trust users put into such system (Adadi & Berrada, 2018; Barredo Arrieta et al., 2020; Martens & Provost, 2014). While first studies used explainability methods like feature importance (De Nadai & Lepri, 2018; Zhao et al., 2018), they are often focused solely on the tabular data and extracted information from the views. Only the work of Liu et al. (2018) uses explanations directly related to the image. Nonetheless, from the paper's description, it is unclear how the XAI method works. Law et al. (2019) called for future research about visual XAI methods and explicitly stated: "Additional research is also needed to visually explain the CNN model. For example extracting discriminate features between higher and lower house prices [...] can potentially bring greater clarity to the model". A practical motivation was reported in 2021 after the start of this thesis. Zillow, a large U.S.-based real estate platform, used algorithms for their new iBuying branch. Their business model was to buy houses for a low price, make small renovations and repairs, and sell them for a higher market price. This method is also known as house flipping. In the heating and suddenly cooling real estate market due to the Covid-19 pandemic, the algorithm failed to adjust to the market, leading to overpriced buying of houses. This took Zillow out of balance, as thousands of bought homes could not be sold or sold profitably, leading to a lay-off of 25% of the hired employees and write-off of 569 million USD (Clark, 2021). This example shows how important explainability and interpretability of algorithms are, especially for online business models. The lack of visual explainability, the lack of a combination of explainability methods for different views, and the call for future research motivated RQ3 about the explainability of multi-view real estate appraisal models.

The developments in image-based real estate appraisal are summarized in Table 2.3. It can be seen that additional views are used to predict the house price using map data (J. Bin, Gardiner, Liu, & Li, 2019) and building plans (Solovev & Pröllochs, 2021). Regarding satellite images, additional authors (Azizi & Rudnytskyi, 2022; Helber, Bischke, Guo, Hees, & Dengel, 2019; Lin et al., 2021; Nouriani & Lemke, 2022; P.-Y. Wang, Chen, Su, Wang, & Huang, 2021) started to analyze their effects.

While until 2019 MVC were de facto the state-of-the-art approach to combine tabular and image data, the trend remains. Between 2019 and 2022 MVC is preferred over MVNN, as 12 of 16 papers use this model, while only four use MVNN (see Table 2.3).

While XAI methods were not regularly implemented before 2019, it seems that, at least for tabular data, these methods are used to explain the behavior of the ML model. We have to highlight the work of Lorenz, Willwersch, Cajias, and Fuerst (2022), giving a good tutorial about XAI for tabular data in real estate appraisal[5]. Also, explainability has been implemented at least three times for image data. However, often visual data are only used to extract categorical features (interior objects or binary indicator of house price) (Kostic & Jevremovic, 2020; Lin et al., 2021), whereas only one research paper uses explainability on the regression outcome (S. Wang et al., 2021). Nonetheless, the authors use an XAI method for classification instead for regression. This might be because most of the XAI methods for image data are specifically built for classification (Lundberg & Lee, 2017; Lundberg et al., 2018; Ribeiro et al., 2016; Selvaraju et al., 2017; Zeiler & Fergus, 2014; Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016). However, methods for image regressions are sparse as only Regression Activation Maps (RAM) and Sliding Window Heatmap exist (Z. Wang & Yang, 2018; Yang et al., 2018). This further motivates research about RQ3 to focus on regression outcomes for visual explanations.

Having motivated RQ1 to RQ3 and the research objective by giving an overview of technological developments for real estate appraisal, the next chapter will introduce details of research methods to be able to answer these questions rigorously.

[5] This work is not included in the table because it does not use any visual data and focuses only on tabular data and the corresponding ML models

Table 2.3.: Paper-Concept-Matrix of related work in image-based real estate appraisal from 2019 until 2022.
Abbreviations: Multi-view kernel learning (MVK), Multi-view concatenation (MVC), Multi-view neural networks (MVNN). Visual explanation: Regression denoted by R, Classification by C. X marks if the feature or category applies to the paper.

Paper	**Data**								**Model**	**XAI**	
	House attributes	**GIS attributes**	**Satellite Image**	**Exterior Image**	**Streetview**	**Interior Image**	**Map**	**Building plan**		**Tabular**	**Visual**
Helber et al. (2019)			X						MVC		
Fu, Jia, Zhang, Li, and Zhang (2019)	X	X			X				MVC	X	
J. Bin et al. (2019)	X	X					X		MVNN	X	
Johnson et al. (2020)	X			X	X				MVC	X	
Naumzik and Feuerriegel (2020)	X					X			MVC	X	
J.Bin, Gardiner, Li, and Liu (2020)	X	X		X					MVC		
Kostic and Jevremovic (2020)	X		X	X	X	X			MVC	X	C
Kang et al. (2020)	X	X			X	X			MVC	X	
C. Lee and Park (2021)	X	X		X					MVNN		
P.-Y. Wang et al. (2021)	X	X	X						MVNN		
S. Wang et al. (2021)	X	X				X			MVC	X	R
Solovev and Pröllochs (2021)	X							X	MVC	X	
Lin et al. (2021)	X	X	X						MVC		C
Potrawa and Tetereva (2022)	X	X		X					MVC	X	
Nouriani and Lemke (2022)	X		X	X		X			MVC		
Azizi and Rudnytskyi (2022)	X		X	X		X			MVNN		
Summary (Total: 16)	15	9	6	7	4	6	1	1	MVC (12) & MVNN (4)	9	3

3. Methodology

This chapter focuses on introducing the different research methods used during this dissertation. Creating theory is an essential process in science (Gregor, 2006). For information systems, Gregor (2006) specifies different types of theories, which have variant aims and assumptions. In general, the author defines theory as descriptions and explanations about the world that improve the current understanding (knowledge), where sometimes theory makes even predictions about the future or is the basis for prescriptive actions to intervene or to design the current state. Thus, a theory is not only the testable propositions of relationships between constructs but is also related to the mindset and mental understanding of a phenomenon in a context and a method to describe how things should be (Gregor, 2006). Overall, theory needs to be generalizable, which describes the ability to abstract the theory to a broader area.

Based on this broad definition of a theory, Gregor (2006) defined five different types of theories. Type 1 theories are 'Theories for Analyzing', which include taxonomies and frameworks capable of describing and analyzing phenomena. This type of theory primarily provides terms, similar to vocabulary in a language, that define and describe an object and its context. Thus, this research type is often the basis for further research because, without Type 1 theories, no phenomena would be spotted. Type 2 theories are 'Theories for Explaining', which could also be described as the most traditional understanding of research. Building on causal relationships, Type 2 theories help to understand a phenomenon by yielding a reason for what is, why, how and where, and when. Type 3 'Theories for Prediction' provide testable propositions of what will happen that are based on the as-is condition. Nonetheless, this type of theory does not provide a reason for the outcome. Type 4 'Theories for Explaining and Predicting', provide causal relationships and testable propositions to predict a future state. So to say, the 'Theories for Explaining and Predicting' is the combination of Type 2 and 3 theories. Lastly, Type 5, 'Theories for Design and Action' are prescriptive and provide knowledge how to do achieve the desired state.

One differentiation between theories can be made by their aim, being either more explanatory or having a predictive focus (Gregor, 2006). In particular, Shmueli and Koppius (2011) differentiated in detail the procedures of explanative and predictive research and deliminated the differences. Explanative research focuses on finding reasons why phenomena occur, while predictive research aims to predict the phenomena as accurately as possible. While a previous understanding was that both aims are based on the other end of a continuum, lastly Hofman et al. (2021) stated that besides having explanative and predictive models, a new type of research, named integrative modeling, is possible, which combines both perspectives. This type may be similar to Type 4, 'Theories for Explaining and Predicting', suggested by Gregor (2006). Hofman et al. explain that, especially for social science, this new type of conducting research may lead to newly derived knowledge because one uses the explanatory model to predict a phenomenon in unseen situations.

This thesis focuses on two research types, predictive theory (Type 3) and theories for analyzing (Type 1). Thus, this chapter first starts with an introduction of ML as a research method in Chapter 3.1 (Predictive Theory) and will continue with details on a taxonomy development in Chapter 3.2 (Theory for Analyzing).

3.1. Machine Learning as a quantitative research method

With ML as a popular method to analyze data, these procedures have also been examined for being used in research. In the past, quantitative data was gathered explicitly for research purposes based on experiments or large surveys. Approximately 80% of the papers use either traditional qualitative or quantitative methods, while the other 20% is related to other research methods like Design Science and computational methods. The latter has been increasingly used in Information Systems and allows the use of data generated from practice (Recker, 2021; Shmueli & Koppius, 2011). In this thesis, Papers 1, 2, 3, 4, and 5 use ML as a quantitative research method.

While Recker (2021) uses the term computational methods for this type of research, Shmueli and Koppius (2011) labels these approaches as predictive research, while Hofman et al. (2021) coin this term predictive modeling. Nonetheless, all the above terms refer essentially to using ML or DL as a new research method in the quantitative field. However, predictive research does not equal quantitative explanatory research, as there are fundamental differences in the goal, focus, constraints, and evaluation (Table 3.1).

Table 3.1.: Characteristics of Explanatory and Predictive Research based on Shmueli and Koppius (2011), extended by the work of Recker (2021) and Shmueli (2010).

Characteristic	Explanatory Research	Predictive Research
Goal	Explanatory models for hypothesis testing	Predictive models for predicting new instances
Focus	Constructs by variables	Variables
Model Building	Minimize bias	Minimize overfit
Constraints	Interpretable model & Support hypothesis testing & Adhere theoretical model	Usage of variables available before prediction
Evaluation	Explanatory power R^2, Significance of coefficients	Predictive power Predictive performance on out-of-sample
Data Generation	Could be Unorganic Data generated for research (e.g., experiment)	Organic heterogeneous & unstructured, large volume, event-based
Generalizability	High match between $\widehat{f(x)}$ & $f(x)$ respectively $\hat{\beta}_i$ & β_i	High match between $\hat{y}$ and y
Causality	By theoretical model and/or data (counterfactual analysis)	Only associative relationship

$\hat{y}$ and $\widehat{f(x)}$ and $\hat{\beta}_i$ denote the *estimated* outcome, function and effect

Explanative modelling

In classical quantitative research, the aim is to test and verify hypotheses deduced from theory. Therefore, explanatory models are used with a focus on transforming constructs into measurable aspects by variables. In addition, usage of an interpretable model (e.g., Linear Regression), statistical tests (e.g., T-Test to verify differences in the mean between two populations), and models strongly designed by theoretical implications from the domain to create causality characterize this research (Shmueli, 2010). The data analysis is designed to minimize bias in the measurements, which is associated with Type 1 and Type 2 errors as the main risk. The results of the analysis are evaluated either by the significance of the statistical results (hypothesis tests or measurements of the coefficients in a quantitative model) and the explanatory power (R^2) of the same one (Shmueli & Koppius, 2011).

Predictive modelling

In contrast, predictive research is mostly not interested in the effects of a variable or construct but instead in predicting new observations. The focus is to gather as many variables as possible that contain a signal for the prediction (associative relationship). The only restriction is that only variables are used that are also available at all times and also before the prediction. With this being the only modeling restriction, complex black box ML models and variables that only are correlated but not causally related to the outcome can be used. There is no specific set of variables that theory demands to use. In addition, also the mathematical relationship can freely be designed. The predictive model is designed to reduce the prediction error, which is the sum of the squared bias and the variance. As the interpretation of coefficients and statistical tests are not necessary, non-parametric ML models are often used due to their high flexibility (James et al., 2013; Shmueli, 2010). However, they come with the disadvantage of being black boxes with little to no possibility of explaining or interpreting the results (Du et al., 2019; Müller, Junglas, vom Brocke, & Debortoli, 2016). The main risk in predictive models is to overfit, a phenomenon that the model works well on a given data set but can not generalize beyond these examples on new observations. Predictive models are evaluated out-of-sample, so the used data in these studies are split into a training and test set, where the training data are used to build the model, and the evaluation is performed on unseen, new observations from the test set, which is a strong contrast to explanatory studies, which do not perform a data split (Shmueli & Koppius, 2011). Furthermore, neither the explained variance of an effect (R^2) nor the significance levels are of interest. The criteria determining if the performance is well are mathematical error measures like accuracy, F1-score for classification, Root Mean Squared Error (RMSE), or Mean Absolute Error (MAE) for regression tasks. To point out the differences on another example, a house price regression is formed (like in Chapter 2.1.1) as:

$$\widehat{f(x)} = \hat{y} = \sum_{i}^{I} \hat{\beta}_i \cdot x_i \tag{3.1}$$

with

$$\epsilon = y - \hat{y} \tag{3.2}$$

where x_i being the i$^{\text{th}}$ characteristic of a house, $\hat{y}$ the predicted house price and $\hat{\beta}_i$ the estimated price effect of the i$^{\text{th}}$ characteristic, with $\widehat{f(x)}$, being the estimated model. While explanatory studies focus on the right-hand side of the equation with questions about the correct measurement and significance of $\hat{\beta}$ and the specification of the formula, predictive studies focus on the left-hand side and the question if $\hat{y}$ is close enough to y (the observed price) and how to lower the error ϵ.

With the aim to have a very low error ϵ, we use, in this thesis, multi-view learning approaches that take at least the form:

$$\widehat{f(x,v)} = \hat{y} \tag{3.3}$$

with x being classical housing attributes coming in a vector format and v being a type of image represented as a tensor to estimate the house price $\hat{y}$. $\hat{f}$ is either a complex multi-view neural network or a combination of multiple ML or DL models. Furthermore, in Paper 2, the house price prediction function $\hat{f}$ does not only have two inputs x, v but three inputs x, v, g, where g is a vector of GIS variables.

Data generation process

In addition, it is noteworthy that besides the different goals and modeling approaches between explanative and predictive modeling reported by Shmueli and Koppius (2011), also the data generation process is different. Recker (2021) reports that unlike in explanatory studies, where the data is generated explicitly for research purposes, computation (predictive) research uses digital trace data. This data is generated by humans and captured and stored by different information systems. However, these data are organic (existing without the research conducted), so they are a byproduct of another process (e.g., house transaction), where oftentimes heterogeneous and unstructured data is used (see multi-view learning), the volume can be large (e.g., Big Data), and the generation is event-based (triggered by real-world actions of humans, e.g., house sale) (Müller et al., 2016; Recker, 2021). However, it is not just the pure volume of the data. Big Data, Data Analysis, and ML for research open up versatile opportunities, as the data include more details in a fine-granular resolution (depth), e.g., not only a house-price index but actual houses and also in the number of variables (breadth), by including many more aspects in comparison to classical surveys or experiments (Müller et al., 2016).

In particular, we use organic data, as the information about the house price, the house characteristics, GIS data, and various image data types like satellite images, exterior and street-view images were not explicitly generated for our study. They result from other activities, e.g., documentation process for house sales transactions or being used in various business applications. For example, the images used are collected by mapping services like Google or Bing Maps to improve navigation. In contrast, e.g., noise data in GIS are used to monitor and analyze the influence of traffic noise on the health of inhabitants of cities. With the usage of many variables (50+), a considerable breadth can be covered. The analyzed datasets have multiple ten thousands observations and thus often represent the whole data population (all houses) in a city or area. Thus, being a house, the smallest 'unit' of analysis resembles a very fine-granular data analysis possibility.

Generalizability of Predictive Research

Lastly, the question remains: what are the differences between explanative and predictive research in generalizability? Explanative research ensures generalizability by measuring the effects correctly. The aim is to get a data sample representative of the underlying population so that with the help of the data, the estimated model $\widehat{f(x)}$ is as close as possible to the unknown true model $f(x)$, resulting in $\hat{\beta}_i$ resembles β_i (Shmueli, 2010). Again the focus is on the right-hand side of Equation 3.1. Predictive models indicate their generalizability through an accurate prediction of new observations (e.g., on the test set) so that it is indicated that $\hat{y}$ resembles y (Shmueli & Koppius, 2011).

To provide generalizability, we follow this procedure and evaluate the predictive performance of the models based on new (unseen) observations by calculating different metrics like RMSE and MAE. Evaluation procedures follow k-fold cross-validation, where repeatedly another train-validation split is used to calculate the robustness (mean and standard deviation) of the predictive performance (e.g., Paper 3, 4). Furthermore, in Papers 1, 2, and 5, a Spatial-out-of-Sample (SooS) for testing was used. Thus, not only are the houses unknown to the algorithm but also the houses come from other locations (subsample of data distribution), again unknown to the algorithm, so the algorithm has to be robust to new observations and locations to outperform the existing baselines. Using the described testing procedures ensures that the results will hold for future data, creating generalizability.

Benefits of Predictive research

After assessing the differences between explanatory and predictive research, one can ask the question, what benefit do predictive models have for research when in a classical sense, they do not test hypotheses, do not measure relationships significantly, and are sometimes not interpretable due to their black box characteristics? Gregor (2006) points out that especially in cases where researchers are not interested in lower-level effects, or concerned about the plausibility of the underlying theoretical model, or causal relationships have not been discovered yet, predictive-driven research can have benefits. Gregor states that: "[t]he discovery of regularities that allow prediction can be of interest if these were unknown before, especially if the theory's predictive power is of considerable practical importance [...]" (p.626). In particular, predictive research can: help to generate new theories, help to develop new measurements, be a tool to compare competing models, improve existing models, measure the distance between theory and practice, and assess the predictability of phenomena (Shmueli & Koppius, 2011). In the role of generating new theories, ML can be a tool that helps to integrate quantitative data (Müller et al., 2016; Shmueli & Koppius, 2011), respectively extracted patterns from data (acting as an automation tool) for qualitative research processes (Berente, Seidel, & Safadi, 2019). As a part of the theory generation process, predictive analytics can be used to create different operationalizations (measurements) of the same construct. Besides the measurements of constructs, predictive research could also compare two different competing theories with respect to their predictive performance, which might help to decide on a theory. ML models can also improve existing explanative research by, e.g., discovering that the assumed linear relationship is, in fact, non-linear and thus contributes to a better formulated explanative model (Shmueli & Koppius, 2011). In addition, by measuring the predictive performance of explanative and predictive methods, predictive analytics can measure the gap between theory and practice and therefore assess the relevance of theory. Lastly, predictive models can assess how well predictable a phenomenon is (Shmueli & Koppius, 2011).

Concerning the benefits of predictive research, first of all, it needs to be recalled that predictive performance is of interest for AVMs (Law et al., 2019; Limsombunchai, 2004; K. Peterson, 1993), making any contribution in this part valuable. In addition, all five quantitative papers implicitly assess the relevance of the classical (single-view) hedonic pricing model by testing it against multi-view learning algorithms. In addition, the predictability of the phenomena house prices is also evaluated. The results of the papers can be used to improve existing models, as shown later on, they identify new data sources (different image types) and modeling strategies (equivalent to the suggested non-linearities in Shmueli and Koppius) to increase the predictive power of hedonic models. Beyond these use cases, driven mainly by ML research by measuring performance, creating baselines and new ML techniques to outperform those, Paper 5 relates the comparing competing theories, although not two theories in a classical sense are compared. Paper 5 focuses on measuring the importance (predictive performance) of two concepts, namely Search and Evaluation qualities. In the understanding of Shmueli and Koppius (2011), no new measures for similar concepts or generating new theories (through ML) are part of this thesis.

Guidelines for Predictive Research

Besides the role and contribution of ML for Information Systems, predictive research is not an out-of-the-box research approach with a guarantee to make interesting contributions. As it is a relatively new method used for Information Systems research (see the beginning of Chapter 3.1) and so the familiarity of researchers with this technology is low, researchers should follow rigorous guidelines. Müller et al. (2016) provide these guidelines, which influence all stages in a research project. In particular, a theoretical triangulation is necessary for the creation of the research question as well as the result interpretation. The framing of the research question and aiming at a theoretical contribution are important factors. Further requirements arise in the data collection phase. As the data is often based on trace data, unlike in experiments and surveys, which are explicitly designed to create data for the research purpose, predictive studies should justify and evaluate the selection of the data in terms of validity and reliability. Also, if applicable, access to the data should be granted. For the data analysis, algorithms used beyond the field of information systems (e.g., in Computer Science) should be considered, as ML is a fast-changing and evolving field, which is necessary to use state-of-the-art technology. Moreover, data analysis steps should be thoroughly documented and published, as well as empirical results should be validated. In the case of interpretation, especially when black box models are used, adding explainability could be a desired step to find more evidence for a theory. In particular, it needs to be ensured that theories are not extended based on volatile correlations (Müller et al., 2016).

All in all, predictive research and ML, although new, are methods that can lead to essential contributions, as it helps to integrate data and knowledge in new ways in social science research (Hofman et al., 2021; Lazer et al., 2009). While the majority of papers in the thesis are linked to predictive research, through the research process, it was discovered that a taxonomy for explainability and interpretability in intelligent systems is necessary to describe the phenomena better. Henceforth, in the next section, research methods for analyzing, in particular for taxonomy development, are introduced.

3.2. Taxonomy development

'Theories for Analyzing' are, after Gregor (2006), the most basic form of theoretical contributions. Research belonging to that category has the aim of describing the current state. Thus, sometimes, descriptive research as a metaphor is used. However, despite the fact that some taxonomies summarize the 'what is', identified relationships could be classificatory, compositional, or associative. Maybe the most prominent example of a 'Theory for Analyzing' are taxonomies in biology, where flora and fauna are separated into species, types, and groups. Typical outputs of these Type 1 theories are classification schemata, frameworks, or taxonomies, being numerously available in Information Systems Research (Gregor, 2006), e.g., a taxonomy for Decision Support Systems (Steiger, 1998). In general, Type 1 theories are especially appropriate “when nothing or very little is known about the phenomenon in question” (Fawcett and Downs (1986) in Gregor, 2006, p.623), and can thus be the basis for future (other) theory types, because identification of a phenomenon, and the description with appropriate vocabulary is possible. In this thesis, Paper 6 is a framework to describe the intelligent system's XAI capabilities.

Taxonomies

Nickerson, Varshney, and Muntermann (2013) define a taxonomy as “systems for grouping objects of interest in a domain based on common characteristics” (p.338). A taxonomy is thereby an n-dimension set, where each dimension consists of k (with $k > 2$) characteristics that are mutually and collectively exclusive. This means that each object has to have exactly one characteristic per dimension, being neither allowed to have multiple characteristics nor to have no characteristic. As there is no exact or best result for taxonomies, evaluation criteria, unlike in predictive or quantitative research, are hard to determine. Nickerson et al. (2013) define a useful taxonomy when it is concise, robust, comprehensive, extendable, and explanatory. Conciseness is characterized by parsimony, given a limited number of dimensions and characteristics. In contrast, robustness defines a taxonomy in which enough dimensions and characteristics are used to differentiate the objects of interest clearly. Linked to this, comprehensiveness is sometimes defined as the utility that a taxonomy includes all characteristics of interest. Other definitions relate to comprehensiveness as the ability to classify all known objects in a domain. By extendable, new occurring dimensions or characteristics should be able to be included in an existing taxonomy because it is likely that the described objects change over time. A taxonomy should not only describe the existing object but also explain its nature. Thus, without knowing all the details of an object, it should be located in the taxonomy, as well as vice versa with the taxonomy, not all details of the objects must be known (Nickerson et al., 2013).

Taxonomy development processes

The development of a taxonomy is often characterized by two phases, namely empirical-to-conceptual and conceptual-to-empirical, where each phase could be the starting point, followed by the other. In an empirical-to-conceptual approach, empirical observations, namely the objects of interest either in practice or research, are gathered, analyzed, and compared so that dimensions and characteristics are identified. Similarly, dimensions and characteristics are deduced from theories and theoretical concepts in a conceptual-to-empirical approach. Then, objects are identified that bear these utilities. For studies creating taxonomies, it is necessary to perform both analyses so that one phase is the revision of the other phase, ensuring that both theoretical and empirical aspects are included to provide a holistic view (Nickerson et al., 2013).

Paper 6 focuses on the empirical-to-conceptual approach by analyzing existing AVMs and their XAI capabilities (Kucklick, 2023). The focus is on the empirical-to-conceptual approach, as this paper builds on the work of Kucklick (2022a), where the conceptual-to-empirical phase was conducted.

Meta characteristics and ending condition

Before beginning a taxonomy development process, meta-characteristic and ending conditions need to be defined. The meta characteristic is necessary to determine the context of the development by clarifying the user group of the taxonomy. Oftentimes, the objects of interest are so complex that unfocused, no clear taxonomy can be developed as dimensions are versatile (Nickerson et al., 2013). Analyzing an object's dimensions by reviewing characteristics without an aim is similar to finding random patterns in quantitative data. Due to that, the development process is more guided by a clear definition of meta characteristics, highlighting the purpose and user group of the taxonomy. In addition to the meta-characteristic, defining a stop criterion for the development process is equally important. There are subjective and objective stop criteria, summarized in the work of Nickerson et al. (2013). Objective ending conditions can be evaluated objectively. These include, for example, that all objects in a representative sample can be mapped in the taxonomy, that no dimensions or characteristics were added or merged in the last development step, or that dimensions are unique and not repeated. The subjective ending conditions are related to the formulated definition of a good taxonomy, being concise, robust, comprehensive, extendable, and explanatory (Nickerson et al., 2013).

Paper 6 uses the meta-characteristics of XAI capabilities of intelligent systems, which should be made descriptive, analyzable, and classifiable for data scientists. It meets the objective ending conditions of being able to review all objects in the representative sample (50 papers) used in Paper 6, while no dimensions and characteristics were added or changed in the last development round. Thus, this taxonomy is comprehensive. Furthermore, it is robust, as six different archetypes were identified in the analyzed sample.

Extended Taxonomy Design Process

Building on the work of Nickerson et al. (2013), Kundisch et al. (2022) proposed a structure to support taxonomy development processes named Extended Taxonomy Design Process (ETDP). The authors noticed that oftentimes, taxonomy development processes lack consistency and transparency. Therefore, they adapted from Design Science Research, methods and processes that ensure both details. The proposed method includes 18 steps, from specifying the observed phenomenon to reporting the taxonomy, including empirical-to-conceptual and conceptual-to-empirical approaches. Paper 6 follows the ETDP, with more details provided in the paper (Kucklick, 2023).

Evaluation

Within the ETDP there are several phases that include testing and validation of the suggested taxonomy (Kundisch et al., 2022). While most steps relate to showing the favored qualities of a taxonomy like robustness or conciseness (Nickerson et al., 2013), the Information Systems discipline always focused on building usable theories for practice. The applicability check suggested by Rosemann and Vessey (2008) is a helpful research method, not just for taxonomies, but for all types of Information Systems research, to demonstrate the strong fit between suggested theory and usability. Dimensions of relevant research are: importance (of the topic), accessibility (understandability and readability of research articles), and applicability (of theories). Among applicability, suitability, or a good fit between satisfying the practitioner's needs to solve his problem and the importance of a problem, are essential concepts that need to be satisfied.

Overall, the applicability check of Rosemann and Vessey (2008) can be seen as an evaluation method of existing research conducted at the end of a project or as stand-alone research. Furthermore, it could also be applied within research life cycles, conducted at the end of a project to start a new one (the improvement of applicability of existing research). Favored methods within the applicability check are to conduct interviews or focus groups with practitioners, where the suitability and accessibility of research can be evaluated.

To further evaluate the applicability of the taxonomy in Paper 6, two focus groups were conducted with data scientists from various industries. They conclude that the proposed taxonomy is suitable to support data scientists in the development process of ML-based systems (Kucklick, 2023).

Generalizability of taxonomies

Going back to the definition of theory from the beginning of Chapter 3, generalizability should be inspected for 'Theories for analyzing'. Due to objective and subjective ending conditions, taxonomies provide generalizability through their development process. One objective ending condition, which is often used, is that the taxonomy can describe all objects in a representative sample, aiming to provide generalizability indicated by that sample. This relates to the subjective ending condition of comprehensiveness that all objects and dimensions of interest are included in the taxonomy. Both do not just define the termination condition but are evaluation criteria indicating when generalization is reached in the development process. Paper 6 follows the described process (Kucklick, 2023).

In this chapter, the research methods used in this dissertation were inspected. The following Chapter discusses the published research papers and their results. Furthermore, the papers are set in relation to each other.

4. Results

In this chapter, the findings are summarized. Therefore, the research papers of Part II are set in relation and context. A short summary of each paper's contributions is given in Chapter 4.2. Afterward, a cross-study generalization (Chapter 4.3) is performed.

4.1. The house of real estate appraisal

A general overview of this thesis is provided in Figure 4.1, called the **House of Real Estate Appraisal**, in which the six research papers are embedded. Naturally, each house needs a foundation. This work is rooted in two areas: The real appraisal domain, including to have a background in the hedonic pricing model, the requirements and stakeholders of this financial application, and incorporated theories of economics (SEC Theory) about product evaluation and its associated costs. The second foundation is multi-view learning, including Computer Vision, Machine, Deep Learning, and XAI. The House of Real Estate Appraisal has two floors, representing two different streams followed in the papers. The first one is the predictive value of data and algorithms. The second one describes the need for XAI and different XAI methods. This builds the structure to provide ML-based real estate appraisal for AVMs or CAMAs, represented as the roof. Papers 1, 2, and 5 focus on different aspects of which data and algorithms improve the predictive performance in appraisal. Papers 3, 4, and 6, located on the second floor, focus on making the suggested solutions from the first floor more explainable using XAI methods. Obviously, a second floor could not exist without a first floor. Thus, the papers on the second floor tend to shortly compare data and algorithms concerning their performance (similar to the first floor) but then deeply inspect their explainability. In particular, Paper 6 builds a framework for data scientists to increase the interpretability and explainability of their designed intelligent systems.

With more details provided in Table 4.1, a connection to the related work analyzed in Chapter 2.5 is drawn. Five of the six papers used ML as a research method (Papers 1-5). All papers focused on multi-view learning by combining tabular data and images.

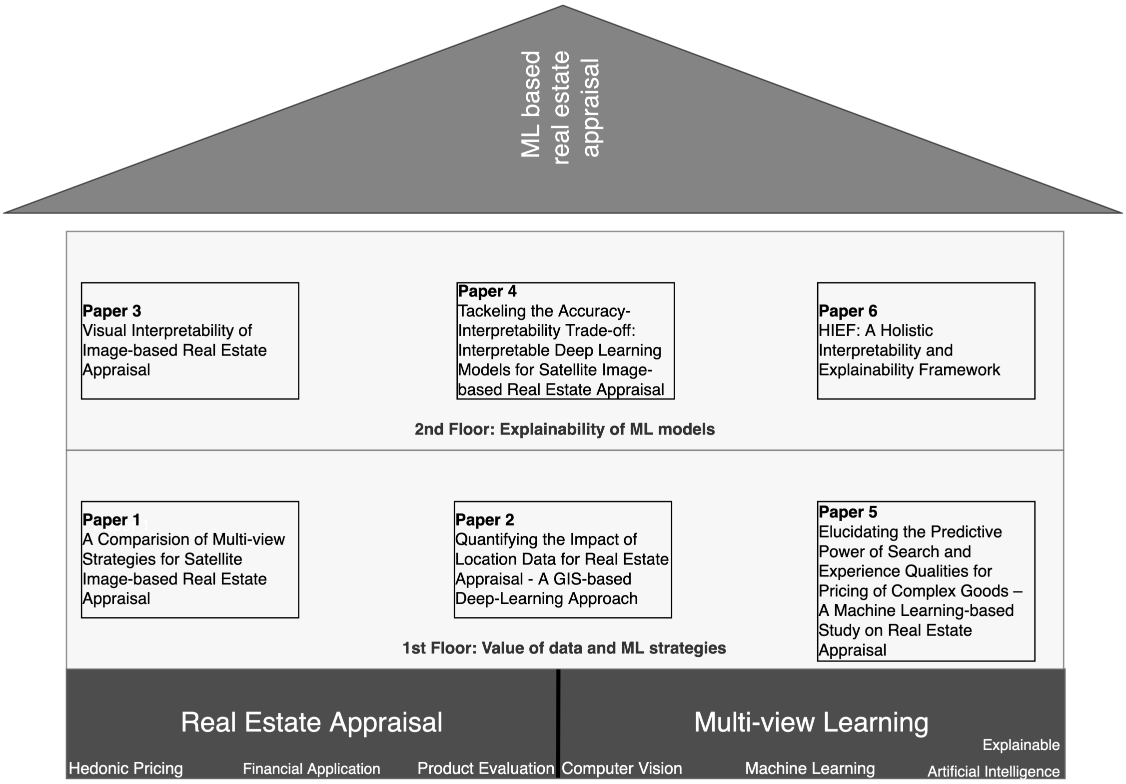


Figure 4.1.: House of Real estate appraisal. This overarching figure visualizes the published paper in this work. Papers on the first floor inspect the value of different data sources and the predictive power of different multi-view learning strategies. Papers on the second floor are built on top of this research and incorporate explainability aspects.

Therefore, all papers used classical house attributes — two of the five papers incorporated in addition also GIS data (e.g., distances to POIs). Three different types of images were analyzed: satellite, exterior, and streetview. Thus, as identified in Chapter 2.5, this work contributes to the related research by measuring the predictive performance of satellite images, which only two out of ten papers did before 2019. In addition, it compares the predictive power of different image types. Moreover, a variety of modeling strategies were examined in this thesis. The focus was on MVC and MVNN. Especially, the latter type of method was used only in three publications until 2019 and only in four out of 16 publications from 2019 to 2022, being not a popular modeling choice. This research could potentially share more details on why and how to build MVNN, discussed in the next chapter. Lastly, XAI methods, especially for visual data, were used only in four out of 26 cases until 2022, each time with substantial limitations (e.g., only being able to explain classification instead of regression decisions). Therefore, this thesis focused on visual XAI methods, extending the understanding of XAI for visual regression tasks. Despite the technical development and use of XAI, the lack of XAI methods in real estate appraisal research motivated a theoretical perspective on this topic. Thus, paper six contributes a framework for interpretability and explainability by building on the lessons learned from the related work and own published research.

After giving a global view on the different papers, the papers are individually introduced in the next chapter with a summary of their results and contribution.

Table 4.1.: Paper-Concept-Matrix of published paper from this thesis
Abbreviations: Multi-view kernel learning (MVK), Multi-view concatenation (MVC), Multi-view neural networks (MVNN). Visual explanation: Regression denoted by R, Classification by C. X marks if the feature or category applies to the paper.

No.	**Paper**	**Data**								**Model**	**XAI**	
		House attributes	**GIS attributes**	**Satellite Image**	**Exterior Image**	**Streetview**	**Interior Image**	**Map**	**Building plan**		**Tabular**	**Visual**
1	Kucklick and Müller (2021)	X		X						*	X	
2	Kucklick et al. (2021)	X	X		X					MVNN		
3	Kucklick (2022b)	X			X					MVC	X	R
4	Kucklick and Müller (2023)	X		X						MVNN	X	R*
5	Kucklick et al. (under review)	X	X		X	X				MVC	X	
	Summary (Total: 5)	5	2	2	3	1	0	0	0	MVC (3) & MVNN (3)	4	2
6	Kucklick (2023)	**Taxonomy**										

*: All types (MVKL, MVC, MVNN) were discussed in this publication.
R*: A new explainability method, called Grad-Ram, was developed to overcome the limitations of RAM applied in (Kucklick, 2022b). Grad-Ram is also applicable in MVNN cases.

4.2. Summary of the papers

Paper 1 compares the performance of different multi-view learning strategies. Therefore we used 32,700 houses from Asheville, NC, where classical house attributes were combined with satellite images. The predictive performance of AVMs was increased by up to 21.5 % when satellite images were included. Three approaches were tested for the multi-view learning strategies: MVK, MVC, and MVNN. MVC performed better than MVK, while MVNN performed best. Nevertheless, a trade-off between accuracy and interpretability was found. Special forms of MVC strategies using inherently interpretable models and hybrid MVNN indicated the possibility of interpreting coefficients. However, although these algorithms were up to 14.1% better than the Linear Regression baseline, they lost performance compared to a fully obfuscated MVNN model. Consequently, two directions of future research emerged: The search for and the comparison with other data sources to examine the performance of multi-view learning algorithms in the context of real estate appraisal or a focus on XAI to increase the interpretability of MVCs and MVNNs. Specifically, the analyzed interpretability aspects were related mainly to tabular (house attributes) data in this paper and therefore need to be extended to the visual data (satellite images).

Paper 2 based on Paper 1 and further elucidated the predictive power of image data. Based on 71,000 houses in Philadelphia, PA, the predictive power between GIS and exterior images in combination with classical housing attributes were analyzed. As a modeling strategy MVNNs were used. Evaluating the RMSE, a combination of exterior images, GIS features and housing attributes performed better than a combination of GIS and housing features, outperforming a combination of exterior image and house attributes (Table 4.3). All models were better than the hedonic housing attributes only baseline. Despite the metrical performance, the data analysis costs were also inspected. GIS data often required a dedicated feature engineering phase, in which variables were designed based on the data (e.g., distance to schools or noise level measurements). This requires often a very good domain understanding and the search for suitable data sources can get cumbersome. In contrast, the CNN working as an automatic feature extractor could use the images without feature engineering and complex preprocessing. However, creating CNNs requires technical expertise and in comparison to classical ML algorithms demands more computational resources and training time due to computational complex models.

Paper 3 inspected the interpretability of multi-view learning algorithms, especially to provide visual explainability. It uses 62,641 homes in Philadelphia, PA, and a combination of exterior images and house attributes. With the provided visual explainability, it advances on Paper 1 and answers the call of Law et al. (2019) for XAI methods that explain which part of an image was vital in the house price prediction. This paper used Regression Activation Maps (RAM) (Yang et al., 2018), one existing explainability method for image regression tasks. RAM had strong architectural limitations concerning the CNN, e.g., no fully connected layers (MLP) could be used after the convolutional layers. Thus, the CNNs architecture needed to be changed. Furthermore, a MVC strategy was chosen, as due to the limitation of RAM, it could not be applied to multi-view neural networks. To provide as much interpretability as possible, a Linear Regression was used as the final model in the MVC. Consequently, the paper identified important patterns in exterior images (e.g., vegetation in the street and front yard positively influences the price) and had local and global transparency by interpreting the coefficients. It confirmed Papers 1 and 2 results' that images bear additional information that could boost the predictive performance. The models in Paper 3 were up to 5.4% better than a Linear Regression baseline.

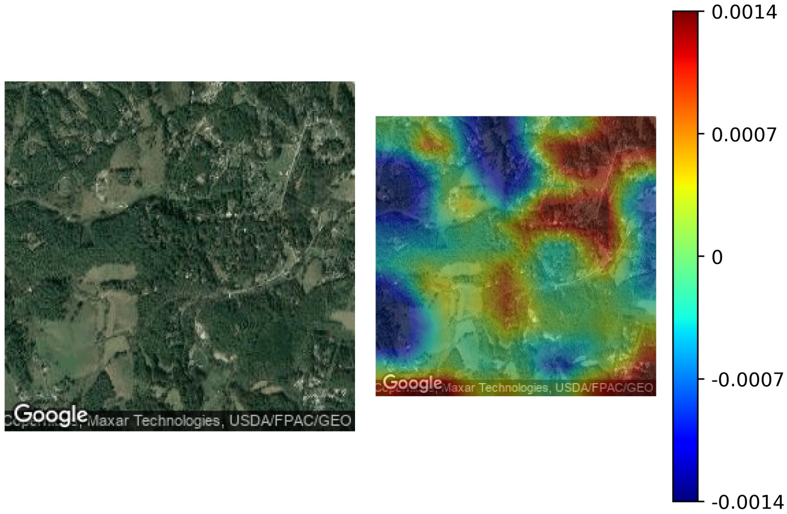


Figure 4.2.: Explanation of Grad-Ram for satellite image on zoom level 14, covering 2400 by 2400m around the house. Appraised real estate is in the center of the image. Red indicates a price increase; blue indicates a price decrease. Changes are measured in standardized log-prices in USD. Original images © Google, Maxar Technologies, USDA/FPAC/GEO, 2022 (Google LLC, 2022).

Paper 4 built on top of the limitations of RAM in Paper 3 and extended this method to Gradient-based Regression Activation Maps (Grad-Ram), with an example explanation visualized in Figure 4.2. This research paper used 32,089 houses in Asheville, NC, combined with differently zoomed satellite images. The new proposed method Grad-Ram can be used to explain MVNN. It can deal with any configuration of layers after the convolutional layers. Unfortunately, Grad-Ram seemed to be not faithful. Thus the usability of this XAI seemed limited. Despite the results of XAI, the paper also reported an improvement in predictive performance by using satellite images and supports the findings of the previous papers. Furthermore, it showed a strategy using multiple XAI methods to generate additional insights (e.g., density information about the neighborhood was extracted from the satellite images). While the explainability could be increased, it was not on par with a Linear Regression. Consequently, an accuracy-interpretability trade-off existed, as the used MVNNs are up to 34% better than the Linear Regression. In terms of a theoretical contribution, the paper provided a socio-technical framework for XAI, connecting the work of Adadi and Berrada (2018), Guidotti et al. (2018) and the 7-gaps model from Martens and Provost (2014) to match user group, aim, and technical requirements into a single holistic view that could be used to deduce requirements for XAI methods.

Paper 5 added another dimension to the discussion of why image data improved the predictive performance of real estate appraisal methods. Paper 1 to 4 have had a technical view of the topic, measured the performance increase, and showed and evaluated different XAI methods. In this context, some reasons have been deduced by manual analysis, which factors of images were important. However, these were not connected to any economic theory. Therefore, Paper 5 introduced the SEC view of product evaluation and measured the performance of Search utilities (variables known prior to the procurement process of the house, e.g., number of bedrooms) and Experience qualities (e.g., noise level or safeness of the neighborhood), which usually need to be experienced by the buyer to incorporate in the decision-making process. Therefore, information from GIS data and images was extracted (feature extraction by CNNs) and classified into Search and Experience qualities. Based on 62,641 houses in Philadelphia, PA, computational ML experiments were conducted. Models were based on a MVC strategy in which different views (Search qualities, Experience qualities, or a combination) were tested. Search qualities alone outperformed Experience qualities. However, when Search and Experience qualities were combined, the predictive performance increased by 15.4% compared to a Search qualities baseline alone. Essential variables were the average price and size in the neighborhood, the size of the house, the physical state, and crime incidents around the house. While size, for example, is an attribute that can be inspected well before the purchase (search quality), the physical state and crime are truly perceived after living in the real estate for a while, representing Experience qualities.

Finally, **Paper 6** used the gathered practical insights from Paper 1 to 5 and indications from related work (not providing a sufficient level of explainability by often neglecting the images) to provide a taxonomy, named Holistic Interpretability and Explainability Framework (HIEF), directed at data scientists. This paper completed the taxonomy development process of Kucklick (2022a) with the empirical-to-conceptual approach and the analysis of 50 AVMs. The authors noticed that many XAI and HCXAI approaches currently rely solely on explainability. Furthermore, the focus is only on an algorithmic perspective (inherently transparent model vs. black box and post-hoc explainability methods). The framework developed extended the currently used views to include aspects about interpretability because it can enhance explainability. Additionally, the framework defined an intelligent system as the combination of data and algorithms and thus opened new dimensions that need to be analyzed. The performed applicability check with data scientists showed that the framework is perceived as helpful. Lastly, based on the framework, intelligent systems published in the real estate appraisal application were categorized, leading to six different archetypes of systems, ranging from highly transparent systems from economics research to obfuscated, very technical ML-based solutions with various approaches in the middle.

4.3. Cross-study generalization

An overall table for a metrical performance (Table 4.3) is created to interlink the results of the previously described papers to generate cross-study generalizability. To compare the results, improvements are measured relative to the baseline. This was a classical hedonic regression model trained only on house attributes in all papers. This algorithm was chosen as it is the used standard in real estate appraisal because it represents the hedonic model well and variables are selected based on the theoretical understanding of a house (Law et al., 2019; Limsombunchai, 2004). Furthermore, it is very interpretable. In addition, we provide metrical details of advanced baselines like a Random Forest or a Neural Network in Table 4.2[1]. We used the advanced baseline models as they have

[1] We reported the results of the advanced baselines that were used during publication. In some cases, through the reviewing process, the results were excluded. Random Forests and Neural Networks are advanced ML algorithms that can account for non-linearity and interaction effects. We used a hyperparameter search for training the Random Forest.

Table 4.2.: Performance evaluation of advanced baselines (Random Forest, Neural Network) on housing characteristics only. Mean MAE and RMSE reported.

Paper No.	Model	MAE	RMSE	Improvement in RMSE compared to baseline
1	Random Forest	38,408	66,268	7.3%
2	Random Forest	33,164	39,174	0.7%
3	Random Forest	17,886	29,651	11.2%
4	Neural Network	36,704	58,676	5.6%

fewer restrictions in terms of linearity and additivity compared to the Linear Regression baseline. We used the RMSE as the metric for comparison, which is a well-used metric for a regression task punishing extreme errors more in comparison to the MAE. While all papers reported the RMSE, some based their evaluation on the MAE. Thus, the improvements in percent from Papers 1 and 2 are calculated based on the reported RMSE and, therefore, can deviate from the text in the paper, which was based on the MAE.

The hereafter described results should be seen rather indicative than evidential. Nevertheless, they help to summarize the gathered results of the thesis by calculating different pivot tables, taking different views on the predictive performance. On average, the multi-view learning algorithms performed 11.7% better than a Linear Regression and 5.9% better than models using ML and housing characteristics only (single view) (Table 4.4). MVKs performed the worst and could not improve the real estate price estimation. MVC improved the predictive power on average by 11.5%, while MVNNs performed best with an average gain of 14.3% (Table 4.4). With respect to the advanced baselines, only the MVNN could consistently outperform the single-view ML models. There are mixed results concerning MVCs. While in Paper 3, they could not perform better than the advanced baseline, in Paper 1, the modeling strategy led to a 6.9% improvement compared to the ML baseline. Concerning the datatypes, in addition to classical housing characteristics, exterior image or GIS data were almost on par with an average of 4.3% and 4.4% better RMSE. Combining the data sources lead to a reduction of prediction error on average by 8.5%. Using, in addition, streetview images increased the predictive performance gains on average to 15.4%, being the best data combination. Satellite images in combination with house attributes increased the predictive power, on average, by 14.5%, being the second best view (Table 4.4). Analyzing the satellite images in detail, the predictive performance decreases by greater zoom levels (more details about the house, fewer details about the neighborhood). Zoom level 14 (2400m by 2400m around the house) performed best by an average performance boost of 23.0%, while Zoom level 16 (600 by 600m around the house) yielded on average a 17.8% increase, and Zoom level 18 (400 by 400m around the house) a 15.5% increase in predictive performance, all results measured for MVNNs (Table 4.4).

Regarding explainability, Table 4.5 summarizes the gathered results. The Linear Regression as an inherently interpretable model provided coefficients that could be used for local and global interpretability. Unfortunately, the analysis and interpretation of visual data was not possible. By default, MVC and MVNN were black box models that neither provided interpretable coefficients for the analysis of structured data nor functionality for analyzing image data. Nevertheless, using different XAI methods improved the explainability of both models. For MVCs, coefficients and visual interpretability could be provided when RAM was used as a visual interpretability method in combination with the suggested modeling strategy in Paper 3 of using a CNN in combination with a Linear Regression. Using Grad-Ram (Paper 4) improved the visual explainability of MVNNs. In addition, when this method was combined with other post-hoc explainability methods (e.g., feature importance), additional insights into the model were generated. To also provide coefficients, a Hybrid MVNN would be needed to be trained (as shown in Paper 1) and could be combined with the suggested Grad-Ram method of Paper 4 to provide coefficients and visual interpretability.

The results from the two paragraphs above can be visualized in relation to each other (Figure 4.3), reflecting on the accuracy-interpretability trade-off. While our results indicated that there is such a trade-off, XAI methods helped to close it. Depending on the strategy, the explainability could be extended, e.g., by using RAM and a model with coefficients in Paper 3. Similarly, a hybrid MVNN in combination with Grad-Ram could be equally interpretable. The MVNN is rated with less interpretability because no coefficients of the model can be used. To complete the figure, we add the single-view (only on house characteristics) trained ML models being used as the advanced baseline model. In general, these models performed similarly to MVCs, but are also black box models, providing no coefficients for interpretability. As demonstrated in Lorenz et al. (2022), XAI methods can help increase these models' explainability. However, they are

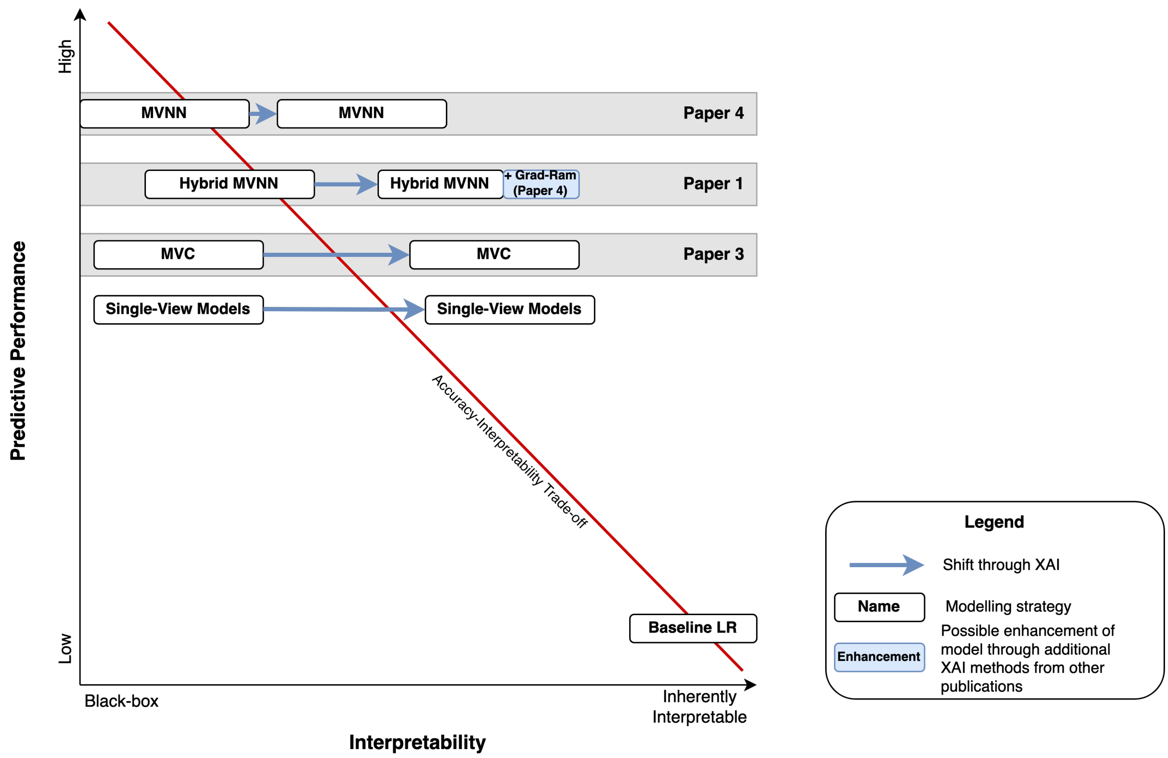


Figure 4.3.: Accuracy-Interpretability Trade-off. Shifts through using XAI are visualized with blue arrows. Note that although XAI methods are applied, interpretability is not on par with the baseline Linear Regression.

rated with a bit higher interpretability because these models and consequently used XAI methods are less complex due to using only a single view. All in all, one has to admit that, while XAI methods helped to close the gap between accuracy and interpretability, the Linear Regression still is the gold standard because the training process and loss surface can be easily understood and confidence intervals and t-tests for significant coefficients are directly available. Thus, the user has no need to rely on XAI methods, which are a model of the used ML model. These can be flawed or have a low fidelity (Adebayo et al., 2018; Tomsett et al., 2020).

Besides the accuracy-interpretability trade-off, Papers 5 and 6 add theoretical contributions from an economic and HCXAI standpoint. Paper 6 extended the understanding of explainability by including data aspects and ideas of interpretability. Paper 5 gave an economic view into the search and pricing process through the SEC theory. Through the transformation of attributes by using ML and GIS, additional attributes that could usually only be experienced after the purchase (e.g., noise level or security aspects) could be measured and incorporated into the pricing process.

Table 4.3.: Comparison of metrical results of published papers. The predictive performance of the best model is reported in RMSE in relative performance to baseline linear regression using classical house characteristics or advanced baseline based on a complex ML model.

Paper No.	**Dataset**	**Datatype**	**Test procedure**	**Multi-view learning strategy**	**Improvement to baseline**	**Improvement to advanced. baseline**
1[2]	Asheville	Satellite Image Zoom Level 16 (600m around house)	SooS			
				MVK	-10.4%	-19.2%
				MVC	13.7%	6.9%
				MVNN (hybrid)	14.1%	7.3%
				MVNN	21.5%	15.3%
2[3]	Philadelphia		SooS	MVNN		
		GIS			4.4%	3.8%
		Exterior Image			3.2%	2.6%
		GIS & Exterior Image			8.5%	8.0%
3	Philadelphia	Exterior Image	5-fold RCV	MVC		
					5.4%	-6.5%
4	Asheville	Satellite Image	5-fold RCV	MVNN		
		Zoom Level 18 (400m around house)		Model by Law et al.	11%	5.8%
		Zoom Level 14 (2400m around house)		Model by Law et al.	11.9%	6.7%
		Zoom Level 18 (400m around house)			20.0%	13.9%
		Zoom Level 14 (2400m around house)			34.0%	25.6%
5	Philadelphia	GIS & Exterior Image & Streetsview Image	5-fold SooS	MVC		
					15.4%	-

RCV: Random Cross-Validation, SooS: Spatial-out-of-Sample,
MVK: Multi-view kernel learning, MVC: Multi-view concatenation, MVNN: Multi-view neural network

2,3: Results reported on MAE deviate from results in paper based on RMSE.

Table 4.4.: Performance evaluation by different dimensions. The mean performance of the best models in the corresponding papers were used.

Groupby	**Mean improvement in RMSE compared to baseline**	**Mean improvement in RMSE compared to adv. baseline**
Algorithm		
MVK	-10.4%	-19.2%
MVC	11.5%	0.2%
MVNN (incl. hybrid)	14.3%	9.9%
Datatype		
Exterior Image	4.3%	-2%
GIS	4.4%	3.8%
GIS & Exterior Image	8.5%	8%
Satellite Image	14.5%	7.8%
GIS & Exterior Image & Streetview Image	15.4%	-
Satellite Image Zoomlevel in MVNNs		
18 (400m around the house)	15.5%	9.9%
16 (600m around the house)	17.8%	11.3%
14 (2400m around the house)	23%	16.2%
Overall	11.7%	5.9%

Table 4.5.: Rating of explainability of developed approaches. Grey rows mark improved explainability through XAI methods suggested in the publications.

Paper No.	Algorithms	Type	Coefficient	Visual Interpretability	Global Interpretability	Local Interpretability	Comment
	LR	Inh. Int.	✓	-	✓	✓	Baseline Hedonic Model
	RF & NN	BB	-	-	XAI	XAI	Advanced baseline Explainability can be improved by using XAI methods see Lorenz et al. (2022)
	MVC	BB	-	-	-	-	Explainability can be improved by XAI methods Example: Paper No. 5
3	MVC	Inh. Int + XAI BB	✓	✓	XAI	XAI	Visual Explainability & Coefficients by modeling strategy enhanced
1,2	MVNN	BB	-	-	-	-	Basic MVNN
4	MVNN	BB	-	✓	XAI	XAI	Enhanced by visual XAI methods and methods explaining tabular data
1	Hybrid MVNN	BB	✓	-	✓	✓	Analyzed in Paper No. 1 could be explanded by visual explainability techniques from Paper No. 4

Inh. Int: Inherently Interpretable Algorithm BB: Black box Algorithm
RF: Random Forest NN: Neuronal Network (Multi-layer perceptron)
✓: Indicator that feature is available, -: Indicator that feature is not available

5. Discussion

This chapter discusses the previously described results regarding the research objective. Recall that the objective is: ***Improving real estate appraisal by leveraging multi-view learning to incorporate multi-modal data.***

Therefore, in the following chapter, we reflect on the research results. Lastly, implications for research in practice are shown in Chapter 5.2.

5.1. Reflection on research results

To progress towards the above-stated objective, we identified two important research areas in Chapter 2.5. First, it is about the performance of AVMs, in particular, which multi-view data to use and which algorithm to train. Second, as AVMs are used in different contexts, e.g., banks, real estate agents, and online platforms, explainability often remains important. Consequently, the explainability of multi-view AVMs needs to be inspected too.

Created by these research areas, three research questions were formulated:

- RQ 1: What predictive power do different data sources (views) hold for real estate appraisal?
- RQ 2: Which strategies in multi-view learning real estate appraisal lead to increased predictive performance?
- RQ 3: To what extent can XAI increase the explainability in multi-view learning real estate appraisal algorithms?

These questions will be answered in Chapters 5.1.1 to 5.1.3. In addition, a conclusion about the objective is given (Chapter 5.1.4).

5.1.1. Predictive power of different data views

With the inclusion of multifarious data sources, including GIS and images for real estate appraisal, the question arose of which predictive value different data sources have. The question gained importance due to the lack of research about satellite images before this thesis started and considering that other data sources like exterior, interior, or streetview images were incorporated, it is relevant for selecting data sources.

The gathered results in Chapter 4.3 indicate that different data sources can improve the predictive performance of real estate appraisal. The averaged effect range was between 4.3% and 15.4% increase in predictive power compared to a Hedonic Linear Regression baseline. Best performed a combination of GIS data, exterior and streetview images, while satellite images were the second best. GIS data and exterior images alone were on par in the performance evaluation. Curiously, the improvement of predictive performance with satellite images increased by 7.5% points when fewer zoomed images (Zoom level 14) were used compared to satellite images that show the house and lot primarily (Zoomlevel 18).

The results support evidence from the related work that the described data sources have predictive value for real estate appraisal. For example, Law et al. (2019) found a 15% increase in performance by using streetview images in addition to the housing characteristics, while satellite images improved the predictive performance by 23%. Similar effect sizes were reported by Bency et al. (2017) about the satellite images. Regarding exterior images, Bessinger and Jacobs (2016) found a 3.6% to 6.8% improvement in predictive power, depending on the ML model used. We can confirm the results from related work with similar effect sizes measured.

In particular, the results with different satellite image zoom levels need further discussion. Bency et al. (2017) tested various zoom levels concerning predictive performance. The authors cascadingly extended their models with a lower zoom level image depicting a higher spatial resolution (more area) and found that this improved the predictive performance. In our studies (Paper 1, 4), we measured the performance gains individually for the different zoom levels. Similar to the results of Bency et al. (2017), less zoomed images improved the predictive performance. Thus, based on our results, it might not be necessary to join multiple zoom levels, but using just one lower zoomed satellite image could result in similar predictive performance without having the need and analysis costs of using multiple images simultaneously. Before discussing possible reasons for this, we briefly describe the identified factors from image data that increased the predictive performance.

The results in our papers indicated that different essential factors for real estate could be extracted from image data. Exterior images accounted for the house's condition and greenery in the lot (Paper 3), both important aspects described in related literature (Donovan & Butry, 2011; Elam & Stigarll, 2012; Johnson et al., 2020; Kostic & Jevremovic, 2020; Quiring, 1987). For example, greenery in the lot increases the house's aesthetical appearance and can have additional effects like reducing heat and improving air quality, or when greenery as a hedge can shield the house against street noise. A well-maintained front yard or good exterior condition indicates that the house and garden are taken care of, highlighting the quality of the house. As specific models were used in Paper 5 for extracting the aesthetics of the neighboring houses and the perceived security of street view images, such implicit factors can also be incorporated into real estate appraisal models. The exterior image in Paper 3 had a similar effect size as the size, or amenities like air conditioning, indicating that the extracted effect is vital for real estate appraisal. On the other hand, satellite images represent location, maybe one of the most critical factors for the house price (Bency et al., 2017; Gröbel & Thomschke, 2018; Hill & Scholz, 2018; Law et al., 2019; Limsombunchai, 2004). The satellite image is much more precise as a dummy variable based on the neighborhood or the zip code. Furthermore, positive aspects like greenery in the neighborhood (e.g., parks), or negative factors like large roads associated with noise and air pollution can also be extracted from satellite images (Kostic & Jevremovic, 2020). In particular, we found that our real estate appraisal model identifies the density (rural or city level) of the house and vegetation aspects as well as land usage (identification of industrial areas) in Paper 4, which is much harder to be accounted for with structured data. Overall, the satellite image was after the size, the second most important variable in Paper 4, and had a considerable influence (coefficient) on the price in Paper 1.

The differences in performance for different zoom levels in the same modeling strategy could have the following reason. Satellite images with a high zoom level (18) depict the area and the house in great detail. However, these factors might already be known from the housing characteristics like (house) size, lot size, or roof style, so there is an overlap between the two views. With a lower zoom level, more details about the neighborhood are captured, leading to details that are only roughly or not captured by the location dummy variables (e.g., neighborhood greenery) - thus leading to a larger performance increase. A similar situation might relate to the exterior image. An overlap between the exterior image and the housing characteristics exists when the housing characteristics report variables like building material (Paper 2, 3) or architectural style.

Previous literature often used new data sources and extracted variables without discussing them from an economic perspective, which we provide here. Factors identified as important from images can be analyzed with the help of two theories, namely hard and soft information from finance (Liberti & Petersen, 2019) and the SEC Theory from economics (Nelson, 1970). Numerical or categorical variables can easily express many variables like size, age, and specific amenities. According to Liberti and Petersen (2019), this hard information is describing facts about an object. Nonetheless, there are, as already identified, factors like the style and aesthetics of a house that also influence the price, but they can not be measured metrically and, thus, are hard to analyze. These factors are called soft information (Liberti & Petersen, 2019). Through the use of ML, especially DL, this information can be extracted, e.g., from unstructured data like images, and made useable for analysis. Law et al. (2019) describe the usage of ML, without considering hard and soft information, as an important factor, because: "[c]ollecting the data required to evaluate urban quality at the city scale is both costly and time consuming. One approach is to cast this as a problem of computer vision" (p.54). Because soft information can be leveraged and incorporated, so real estate appraisal models can make more accurate predictions (Paper 1-4). In addition, it bears after Shmueli and Koppius (2011) also the possibility to use ML as a tool to extract new concepts (one suggested use case) so that theories of real estate appraisal (hedonic pricing model) can be extended in the future. Setting the hedonic pricing model in a greater context beyond quantifying new concepts, the search for information can significantly influence markets and the market participant's behavior, viewed from the SEC theory (Akerlof, 1978; Nelson, 1970). In particular, this theory focuses not on theoretical content (inclusion of important variables) but on the information-sharing process. Houses, as complex goods, have information that can be acquired before the purchase, like size and age (Search qualities), but also many attributes that need to be experienced first (after the purchase) so that information is available, e.g., safeness of neighborhood or condition and quality of parts of a house. This information is hard to acquire before the purchase, and as already mentioned, information asymmetries between principal and agent are increased by online search. Using technologies like GIS, CV, and ML enables algorithms to measure and transform Experience qualities to Search qualities (Klein, 1998), as shown in Paper 5.

Despite the predictive performance, extracted features, and economic theories and implications, technical aspects need to be discussed regarding the data source. Two different types of data are used - namely, GIS and image data. While images are unstructured data, the raw GIS data are semi-structured. They often come in shape files because the objects displayed are points, lines, or planes in relation to one another. The shapefile is thus often a table, with each row an object and the relation to other objects are made by the coordinates. Consequently, the data need preprocessing to be usable (being in first normal form) for any ML algorithm. Such preprocessing needs advanced (domain) expertise by the researcher or data scientists because sometimes different preprocessing can be executed on the same data (e.g., calculating the distance to POIs or conducting a buffer analysis and counting POIs in an area), but one makes more sense than the other. Also, before the analysis, knowledge is needed about which data can be found and which spatial data is relevant. When data is preprocessed, it can be stored in a structured format (e.g., table) and used by various ML algorithms, ranging from Linear Regression to dense neural networks. For example, very little computational complexity is required when a Linear Regression is used. In contrast, image data need no to little preprocessing (e.g., just normalizing) and often can be gathered automatically by querying an API. On the other hand, knowledge extraction is technically very challenging as methods like CNNs need to be mastered. In particular, these impose large hardware requirements like using a GPU. Consequently, while GIS data need more contentwise expertise, image data need technical (DL) expertise to be used in real estate appraisal (discussed in Paper 2).

Reflecting the use of different data sources in multi-view learning real estate appraisal, even more data sources will be used in the future. While related work already extended the views to include, for example, map or building plan data (J. Bin et al., 2019; Solovev & Pröllochs, 2021), also currently published studies state that additional image data and data from IoT-like mobility patterns include valuable information (Naser et al., 2021; Wei et al., 2022). The included factors often relate to aesthetics and style. Consequently, comparing and combining different data sources will remain an important field in the future.

In a nutshell, GIS and image data have improved the predictive performance of real estate appraisal models, where a combination of GIS data, exterior images and streetview images yielded the best performance, but could almost be matched by using satellite images. Most likely, the performance boost comes from including previously omitted soft factors like aesthetics. In addition to content-wise identification of factors, using technology, Experience qualities were transformed into Search qualities, reducing information asymmetries and being ready to be included in the ex-ante house price analysis.

Technically speaking, while GIS data require more preprocessing, image data are computationally more expensive to analyze through the use of CNNs. While the data used is just one factor, we should discuss the influence of different modeling strategies on the predictive value.

5.1.2. Predictive performance of multi-view learning strategies

Within related work, multiple strategies for multi-view learning exist, while mainly MVCs were used. You et al. (2017), Liu et al. (2018) and Law et al. (2019) introduced MVNN for real estate appraisal. With our publications making up for 30% of the research papers using and analyzing MVNNs in appraisal, we contribute to the use and understanding of this strategy in this domain. In related work, often, just one strategy is applied within one paper, mainly to incorporate additional data (c.f. Table 2.2, 2.3). Thus, the question arises about which modeling strategies yield which performance. Throughout this thesis, we tested three different multi-view learning strategies and dedicated a single paper (Paper 1) to comparing the models. Overall, MVNN performed with an average performance increase of 14.3% better than MVCs with an average performance increase of 11.5%, both models compared to the hedonic regression baseline. MVK did not outperform the baseline (Table 4.4).

One possible reason for the under-par performance of MVKs is that, in theory, the errors of the one view should be canceled out with the errors of the other view (C. Xu et al., 2013). Consequently, one implicit assumption is that the different kernels perform equally in predictive power. If one kernel performs significantly better than the other, errors can not be canceled out, leading to a performance worse than the baseline (see Paper 1).

One advantage of MVC like MVKs is that the complexity of the multi-view learning case is split into different views, where separate models can be trained. This leads to a lower resource complexity. Despite the better performance of MVCs compared to MVKs, one technical challenge is the 'Curse of Dimensionality' (C. Xu et al., 2013). For example, when the output of the hidden units of a CNN are used as extracted features in the final model, the number of variables can get very high (multiple thousands) due to the many neurons or convolutional filters in the CNN. This can lead to the so-called 'Curse of Dimensionality', a situation where there are many or in an extreme case more features than there are observations. Overall, this leads to a high risk of overfitting (low generalizability). One example with many features is the work of Bency et al.

(2017), where multiple thousand hidden features were extracted from multiple satellite images with different zoom levels. One way the authors dealt with the 'Curse of Dimensionality' was the choice of a robust ML algorithm (Random Forests), which limits the dimensions through the randomized choice for the feature split in each tree. Despite the technical details, the performance increases from MVCs differed from being worse than a single-view ML model to being 6.9% better (c.f. Table 4.3). One explanation for the performance differences is the model architecture of MVCs itself. While in Paper 1, a Random Forest was used as the last model of this strategy, in Paper 3, a Linear Regression was selected to maintain explainability of the coefficients. In the latter case, this seems to have cost much performance, as the single-view ML model outperformed the single-view hedonic Linear Regression by 11.2% (Table 4.2). Consequently, there are a lot of non-linearities and interaction effects in the data, which are captured by the Random Forest per default, leading to the performance increase. The Linear Regression's assumptions about linearity and additivity of effects are too strong to fit the data optimally. However, following the argumentation of Rudin (2019), with a careful feature selection, creation, and transformation, inherently interpretable models like the Linear Regression could be equally performed than other ML algorithms.

MVNNs may be the most complex solution, as they are not plain vanilla neuronal networks but have to be individually designed and coded. Furthermore, despite the complexity of creating, hardware requirements can be more extensive because of the model in model design, where multiple branches in the neuronal networks are trained simultaneously. In contrast, in MVKs and MVCs the individual models can be trained separately. In addition, it is reported that the optimization can be very challenging (Bency et al., 2017; Kostic & Jevremovic, 2020). Nonetheless, they have a strong advantage. MVCs require that a meaningful feature is extracted from a view (e.g., image) so that it can be harmonized with other views, creating a new feature space. Therefore, an algorithm and or at least a dataset for creating a feature extraction model need to exist, which is often a problem according to Law et al. (2019). The researchers described this as a chicken-egg scenario because these specialized datasets and models often do not exist and thus must be created first. For example, Poursaeed et al. (2018) labeled multiple thousand images into different luxury levels so that the luxury level information could be extracted from real estate interior images and incorporated in the price estimation. The strength of MVNNs is that they create a latent subspace where the features are combined. Thus, only the overall target (e.g., house price) needs to be known, but no additional labels need to be provided. In such a situation, the algorithm, as it is simultaneously aware of structured data (house characteristics) and image data, can better learn required information regarding a potential overlap, leading to better overall performance.

Taking a step back from the different multi-view learning strategies, one has to admit that some performance can be gained by using ML models (advanced baselines) on the single view of housing characteristics. While, on average, multi-view learning strategies performed better than these models, these solutions are easier to implement and bring most certainly enhancements in predictive performance. Thus, the first step in improving AVMs should be to use ML-based single-view models. If performance is the prime directive for an application, then multi-view learning, especially multi-view neuronal networks, should be selected, as this was the only strategy consistently outperforming the advanced baseline models. While also MVCs can create an advantage concerning predictive power, the used models within the strategy have a considerable influence.

To sum it up, MVNNs perform better than MVCs and perform better than MVKs. However, MVNNs are complex to design and train, while MVCs have fewer technical challenges. Although predictive performance is an important factor, the tested strategies have different levels of explainability, as discussed next.

5.1.3. Explainability of multi-view learning real estate appraisal algorithms

Multi-view learning algorithms, as discussed earlier, can get complex. The inner decision-making gets obfuscated, especially when CV and, therefore, CNNs are used. Consequently, the models are at least limitedly interpretable, which contradicts the basic concept of a hedonic pricing model, where the interpretability and explainability of the algorithm are vital functionalities. To increase the explainability of multi-view learning algorithms, we investigated the use of post-hoc explainability methods. In particular, Law et al. (2019) called explicitly for methods that can explain image data. Our research indicated that only two methods are available for regression tasks, namely Sliding Window Heatmap (SWH) (Yang et al., 2018) and Regression Activation Maps (RAM) (Z. Wang & Yang, 2018). Unfortunately, the hyperparameters of SWH strongly influence the explainability (Bansal, Agarwal, & Nguyen, 2020; Du et al., 2019), making it hard to get consistent explanations, leaving RAM as a starting point.

To provide explainability for tabular data and images, we showed in Paper 3 that knowledge about both data types could be gathered by combining MVC as a modeling strategy with a Linear Regression and RAM as an XAI method to explain the CNN.

Nevertheless, RAM opposed substantial limitations on the CNN, e.g., no fully-connected layers were used after the global pooling layer. To better answer the call from Law et al. (2019) and also to be able to provide explainability to MVNN, we developed Grad-Ram. Grad-Ram, the adaptation of Grad-Cam (Selvaraju et al., 2017) to regression tasks, is applicable to MVNN (Paper 4). Various authors suggested different tests for post-hoc explainability methods (Adebayo et al., 2018; Alvarez-Melis & Jaakkola, 2018; Tomsett et al., 2020). While Grad-Ram provided helpful heatmap overlays, in the fidelity tests, weaknesses were revealed. Wheras this method is sensitive to the data and to the model, two critical aspects for a post-hoc explainability method, the effect sizes of the different areas could not be measured precisely, which was indicated by a low AOPC and low faithfulness score. Thus, it remains unclear if highlighted areas in the images have the estimated effect size.

The last two chapters examined the performance differences between different data sources and modeling strategies. The used models are black boxes leaving their inner decision-making process intransparent. While post-hoc explainability methods can help explain the model's decisions and can be used to identify important variables, the transparency is not on par with a Linear Regression building the gold standard (see Figure 4.3). Thus, we can conclude that there is a trade-off between accuracy and interpretability, at least for multi-view learning algorithms. While Rudin (2019) presented evidence that no accuracy-interpretability trade-off exists, at least for structured data, we concur with Adadi and Berrada; Asatiani et al.; Gunning et al. (2018; 2021; 2019), that such a trade-off exists. Our research indicates that this holds true for multi-view real estate appraisal. The trade-off occurs from using DL methods, like CNNs, which perform very well on image data, but are black boxes, making the overall intelligent systems intransparent. Different implications for explainability can be deduced depending on the chosen multi-view learning strategy. MVNNs are harder to explain because post-hoc explainability methods must deal with multiple views and data types. Barredo Arrieta et al. (2020) see great potential in XAI for models that fuse information (multi-view learning), as information about the relation of data and the fusion can identify new important data sources and thus enhance the existing knowledge. Nonetheless, the authors are concerned that information fusion can also violate private information rights, even though carefully designed and checked by XAI. While there are currently no post-hoc explainability methods that can natively deal with images and tabular data simultaneously, we suggested in Paper 4 the approach of using multiple specialized post-hoc explainability methods for explaining the structured data (housing characteristics), as well as explanations for image data individually. While often standard software is available, e.g., for feature importance or PDP, the software needs to be extended to handle multiple inputs. One way to provide coefficients for tabular data and image analysis boosted by post-hoc explainability methods is a combination of hybrid-MVNN (Paper 1) and Grad-Ram (Paper 4). The advantage of a hybrid-MVNN is that because they do not use a fully-connected layer after the fusion of the different input branches and no fully connected layers in the tabular data branch, they might be able to provide the best mix between accuracy and interpretability. In contrast to MVNNs, the model complexity can be broken down in MVKs and MVCs, as multiple models are used. Thus, standard software for explainability can be used. Challenges arise in selecting extracted features and the availability of models and data for feature extraction.

Despite these technical details, it is necessary to reflect on the potential of XAI methods regarding the development process and the domain. As Adadi and Berrada (2018) described that providing XAI is a challenging task, this thesis also holds theoretical contributions for developing post-hoc explainability methods. The 7-gaps model (Martens & Provost, 2014), user groups and their aims (Adadi & Berrada, 2018), and classes of XAI concepts and methods (Guidotti et al., 2018) were already identified as important aspects in related work. Nonetheless, an overarching model combing these three aspects was missing. One could help design and develop suitable XAI methods by matching all three dimensions through such a model. In Paper 4, Figure 1, we provide an integrated view of the topic. For example, from the framework and in the paper, it was deduced that real estate agents need to understand the intelligent system (real estate appraisal) to use it as a decision aid (relating to the 7-Gaps model). By using suitable post-hoc explainability techniques, the real estate agent's mental model and the gap to the intelligent system can be reduced (see Martens and Provost (2014)). Therefore, the agent, as the user role manager (see Adadi and Berrada (2018)), needs control of the algorithm. To ensure that right and lawful features are used, using feature importance (see the classification of Guidotti et al. (2018)) is one suitable method choice.

In addition to the proposed framework from Paper 4, HCXAI introduced methods to include the user in the design process for an XAI method. While such processes seem to be helpful, one strong assumption is that a black box model is already in place. These methods are often used because no feature engineering is necessary, making it possible to provide a solution with minimal effort (Liao & Varshney, 2021). Nevertheless, inherently interpretable models with a knowledgeable feature extraction and engineering process could be equally performant (Rudin, 2019). To enable a discussion about the design of intelligent systems, Paper 6 provides a taxonomy to be able to identify and classify XAI aspects in intelligent systems. In contrast to existing papers in HCXAI (Belle & Papantonis, 2021; Förster et al., 2020; Mohseni et al., 2021) not only aspects about explainability, but interpretability dimensions after Lipton (2018) were integrated in the framework. Furthermore, as an intelligent system is not just the algorithm but also the data used, the framework integrates dimension accounting for aspects of the dataset like complexity (number of variables), information type (hard and soft information), or the understandability of the features. This extends the current view of XAI systems to be holistic. We evaluated the paper with data scientists by an applicability check (Rosemann & Vessey, 2008). The proposed taxonomy was highly appreciated because dimensions important for providing explainability, e.g., the data structure, which were unrecgonized before, were identified by the participants.

Concerning real estate appraisal, the usage of XAI could uncover multiple potentials. First of all, only a few related papers focus on providing full explainability (tabular and visual) for real estate appraisal (see Table 2.2, 2.3). Therefore, advancing usage of XAI might uncover new factors important in the appraisal extracted from images. Furthermore, the usage of AVM might be extended. As a recent survey from National Association of Realtors (2022) indicated, AVMs are the least popular choice in making a price estimated by the appraiser. One possible reason could be that the appraiser's mental model is not aligned with the intelligent system (AVMs). By providing helpful explanations for the decisions of the AVMs, this gap could be closed, leading to more trust and usage after Martens and Provost (2014). Consequently, this might lead to more consistent and less biased price estimates, which are required in the industry (Sing et al., 2021). In addition, by using XAI to make more explainable systems, an 'answer first, explanation second' attitude could be mitigated (Nussberger et al., 2022). Moreover, explainability methods could provide the desired level of transparency for the high-stake or scarce resource decision, making the intelligent system accepted by a broader range of user groups (Adadi & Berrada, 2018; Guidotti et al., 2018; Nussberger et al., 2022).

Concluding, multi-view learning strategies often are related to intransprency, mainly because very advanced ML or DL models are used, limiting the explainability. We researched different ways to enhance the explainability of these algorithms, either by an intelligent combination of modeling strategy and post-hoc explainability methods or by combining multiple post-hoc explainability methods. With the two suggested theoretical implications, the socio-technical view on XAI combining the 7-gaps model, user requirements, and techniques in Paper 4, as well as the taxonomy in Paper 6, helps (a) data scientists in the directed design of XAI methods and (b) in the discussion and evaluation of intelligent systems and their XAI capabilities. The results are more specifically discussed in the real estate appraisal context in the next chapter.

5.1.4. Improvement of real estate appraisal models

Overall, many stakeholders rely on a price estimate for a house, including buyers and sellers, real estate agents, financial lenders, and municipalities (Law et al., 2019; W. McCluskey et al., 1997; S. Peterson & Flanagan, 2009; Poursaeed et al., 2018; Sing et al., 2021). Through the advancements in CAMA and AVMs, the price estimation is performed by an algorithm, reducing the overall workload and giving concise estimates (Koch et al., 2019; Sing et al., 2021). The predictive accuracy is one desideratum of appraisal algorithms (Law et al., 2019; S. Peterson & Flanagan, 2009). By using multi-view learning strategies, multiple data sources like housing characteristics, images, and GIS data can be incorporated into the pricing algorithm (Papers 1, 2, 3, 4, 5). This offers the opportunity to integrate factors previously excluded from the analysis, like soft information (Paper 1, 2, 3, 4), as these are hard to express in a structured format, as well as Experience qualities from the good, like noise levels or quality aspects of the house (Paper 5). The additional data sources, combined with advanced modeling strategies (MVCs, MVNNs), increased AVMs' predictive accuracy.

Nonetheless, the above-described models are obfuscated. Thus, additional explainability needs to be provided. We, therefore, investigated the role of modeling technique and post-hoc explainability methods (Paper 3) or the use of multiple post-hoc explainability methods (Paper 4), including the adaptation of Grad-Cam for regression tasks called Grad-Ram. Despite the technical contributions, two frameworks were developed to help in the design process, one related to matching user groups - aims and post-hoc methods (Paper 4) and one directed at the algorithm and data aspects of interpretability and explainability of systems (Paper 6). By this means, we hope to overcome the hurdles of having a large gap between the intelligent system and the user and having non-specific post-hoc explainability techniques.

In particular, we can give the following recommendations: For applications within real estate appraisal, like calculating the property tax based on the real estate value, interpretability might be absolutely necessary. The algorithms must always be explainable to users (employees of the municipalities or taxpayers). Thus, a Linear Regression seems to be the right choice. Similarly, this might also hold for financial lenders, as the GDPR binds them to use inherently interpretable algorithms.

For use cases where only the predictive performance is substantial, MVNNs are a promising modeling strategy, as they provide the best performance. Nevertheless, skilled data scientists are necessary to create such a solution. An exemplary use could be for a first price estimate on online real estate websites like Zillow.

Real estate agents and maybe also online real estate platforms as a new feature might need to have reasons for the pricing decisions to get a better domain understanding, be in control of the algorithm, or provide justification to the end user. Therefore, we suggest using the trade-off between explainability and interpretability in form of hybrid MVNNs or MVCs. These algorithms performed better than the single-view ML model and can provide explainability for tabular input data in the form of coefficients and heatmaps produced by XAI methods for image data. In particular, multi-view learning here has two great potentials: (a) to use additional variables like style and aesthetics to form a better model including all relevant factors and (b) to use transformed Experience qualities, reducing the information asymmetries between principal and agent.

We have improved AVMs by using additional data sources like images, extended their explainability by customized modeling strategies and new XAI methods like Grad-Ram, and provided help in designing the intelligent system by a socio-technical framework and a taxonomy of XAI capabilities.

5.2. Implications

This thesis advances the state-of-the-art in multi-view learning-based real estate appraisal. The following implications arise for research and practice.

5.2.1. Implications for research

Implications for research can be grouped into three categories.

First, the techniques used to combine multi-modal data can be used to incorporate either the extracted information or the real estate price prediction into other theoretical models. (A) For example, fellow researchers can use multi-view learning to identify new factors for real estate economics, e.g., new variables extracted from unstructured (image or text) data, non-linear relationships between hard and soft information, or the connection to other economic concepts (e.g., Experience qualities). For example, the gathered knowledge could improve the hedonic pricing model. (B) The ability to extract Experience qualities from data opens up the possibility of better understanding product selection for different goods and the value of Search, Experience, and Credence qualities by combining ML as the underlying research methods for extraction and economic theories like the SEC framework. (C) This thesis improves the precision of real estate price estimates, which are the bases for other economic analyses, e.g., house price index or economic development. Other researchers could use the implemented multi-view AVMs to benefit from the more accurate price estimates in their economic models.

Second, specifically Grad-Ram as an explanation method advances the state-of-the-art explainability methods for visual regression tasks, as currently only two methods, namely SWH and RAM, are available, which have substantial limitations. By applying Grad-Ram to their research, others can identify essential factors from their image regression, strengthening the theoretical understanding.

Third, as we have shown that the combination of multi-model data includes previously uncovered information, researchers can now focus on exploring the created information subspace to understand the relation of different data sources better.

In general, for future research, the gathered procedures, methods, and knowledge from the use case of real estate appraisal could be applied to similar applications like the pricing of used cars, used goods or the price prediction of financial values like stocks (see Figure 5.1). In the three listed examples, (a) always a price (regression) is performed equal to appraisal, and (b) the data structure consists of structured data (tabular data) and unstructured data (images or texts). Thus, a better understanding of hard and soft information, the influence of Search and Experience qualities, or identifying additional variables using XAI could be detected.

For example, in calculating car prices, the power, number of seats, and fuel efficiency are hard information after Liberti and Petersen (2019). Nevertheless, the aesthetics and conditions could also significantly impact the price, which could be extracted from the car's images (exterior images, front view, details, and interior images). Currently emerging research (Tsagris & Fafalios, 2022) about used car prices focuses on the tabular data, leaving the question unanswered if image data can improve the used car price prediction. Furthermore, cars are also complex goods as they possess Search (facts about the car) and Experience qualities (ride quality, dimensions, size, seating position, quality of materials, etc.). Additional insights about the SEC theory could emerge from future research in this field. Closely related is the price prediction of damages on leasing car returns published by Jameel, Arif, Hintsches, and Schmidt-Thieme (2021), who already make use of the combination of structured and unstructured data. As a generalization, the pricing aid of used goods on online platforms could benefit from using structured data and images (e.g., the combination of storage capacity, production year, smartphone battery capacity, and the current condition extracted from images). As such research emerged (Fathalla, Salah, Li, Li, & Francesco, 2020; L. Han et al., 2019), questions about the modeling strategies arise as only MVNN are tested up to now. In addition, the necessity of XAI methods should be inspected too. For financial applications, as suggested by Caron and Müller (2020), structured data about a company and the stock market (e.g., number of sold shares, trade volume, return of investment, and capital) can be used in combination with texts (e.g., financial news or corporate reports) to predict the price of shares. Beyond the price prediction, multiple use cases range from poverty prediction (Sheehan et al., 2019) to process mining (Oberdorf, Schaschek, Stein, & Flath, 2021) using multi-view learning strategies. This work could inspire these research fields to analyze further the used modeling strategies or apply XAI methods better to understand the used decision system and the real world, closing gaps two to seven in the 7-gaps model.

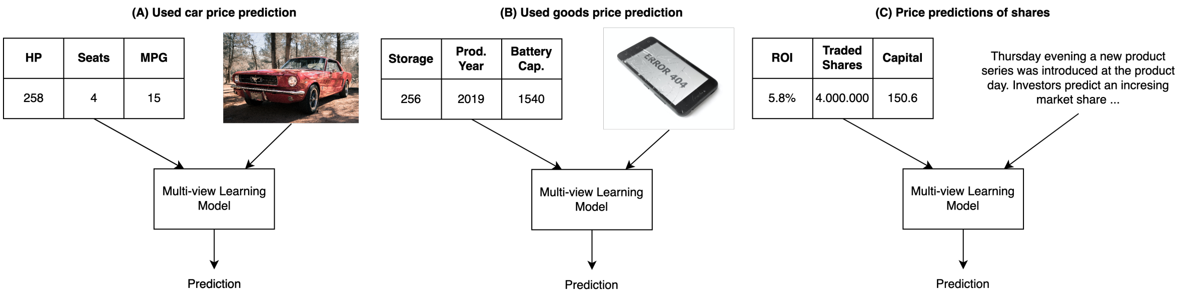


Figure 5.1.: Future applications of multi-view learning, which have a similar data structure. Used cars, used goods, and stock price predictions have a data structure similar to real estate appraisal for which multi-view learning could improve the predictive power.

5.2.2. Implications for practice

A thoughtful decision in practice needs to be made if multi-view learning should be applied to the different use cases of real estate appraisal. The trade-off between predictive accuracy and explainability drives this decision. Our results indicated that the inspected multi-view learning strategies improve the predictive performance by up to 34% and, on average, by 11.7%. While this improvement helps to increase the quality of AVMs, the used methods are black boxes. Although different XAI methods can be applied to increase the explainability, these are not on par with a hedonic linear regression model, which can still be considered as a gold standard regarding transparency. HCXAI procedures should guide the selection process, e.g., the suggested workflow of Mohseni et al. (2021), in which the user group is intensely involved in the development of the ML-based systems, especially in the design of the XAI methods. The 7-Gaps model of Martens and Provost (2014) should be used to understand the mindset and requirements of the user group, and in combination with the HCXAI process, the gap between the mental model and the algorithm should be minimized. In particular, before creating any black box model, the taxonomy of Kucklick (2023) could guide the developer (data scientist) to reevaluate the algorithmic design choices. This could also prevent the 'answers first, inspect later' attitude, identified by Nussberger et al. (2022), describing situations where users were willing to trade interpretability for accuracy, although the user stated prior that interpretability was necessary for this application.

Different contexts in real estate appraisal should influence the choice of whether a multi-view learning application is used. For example, providing just a first price estimate in an online setting, like Zillow's Zestimate®, might be an appropriate use case because users prefer precision over explanation. Other use cases, like the loan application process or property taxation, might require the most transparency. While the former is regulated in some regions by law (e.g., GDPR), the latter might be kept inherently transparent because of the desire of users. As Nussberger et al. (2022) reported, in cases of high-stakes and scarce resources or an algorithm as a gatekeeper function, humans want to have a transparent intelligent system to monitor a fair resource allocation.

Considering data sources, real estate appraisal applications should be extended beyond the currently used housing characteristics to incorporate a plurality of data sources like GIS or image data, as they can improve the predictive performance, one important criterion named for AVMs. In the realm of the 7-Gaps model, by a better algorithm, the gap between the real world and the user's mental model can be closed, providing a better understanding of the domain and the world. Nonetheless, the level of algorithmic complexity in using these data sources should always be considered, e.g., that image data might require intransparent black box models (e.g., CNNs), reducing the explainability of the real estate application.

In practice, information asymmetries between the principal (buyer) and agent (seller) can be dangerous, as they can lead to market failure. Adding additional views through different data sources also helped to incorporate Experience qualities into the algorithmic decision. This can reduce asymmetries and thus strengthen the market and prevent market failure. Especially this attribute could be attractive for online real estate platforms, as the information barrier of only relying on displayed information makes it hard to evaluate a product. Using ML and GIS, the information barrier can be shifted as experience qualities can be made measurable and thus searchable (available before the purchase).

Overall, practitioners should use XAI methods to provide some necessary explainability. Previously, when black box real estate appraisal models were used, XAI methods were not necessarily used. In particular, explainability methods for image data should be applied as just four out of 26 multi-view learning appraisal solutions published between 2012 and 2022 used such methods, which is aligned with the call for more visual explanation from Law et al. (2019).

Beyond the technical requirements, this thesis might encourage different practitioners to embrace the use of AVMs. As currently reported (National Association of Realtors, 2022), appraisers do not use AVMs. One possible reason is the missing understanding of such algorithmic systems and, thus, a large gap between the mental model and the decision system. This gap might be closed with specific explanations and different XAI methods. This furthermore holds the potential to make appraisals more consistent in the future by relying on algorithms (Sing et al., 2021).

In an overarching view beyond real estate appraisal, the gathered results of how to combine different data sources (multi-view learning) in the context of price prediction and the SEC theory could be of interest for further applications like the pricing of used goods, which has many similarities to real estate appraisal (see Chapter 5.2.1). Thus, the algorithms, modeling strategies, and inspected XAI method could appeal to data scientists in this area.

6. Closure

In this chapter, we first describe the limitations. Finally, we give a short outlook.

6.1. Limitations

Our research does not come without limitations. First of all, while the summarized performance in Chapter 4.3 should be seen as evidential, one analysis with different models and strategies and all data sources, including newly available data sources from IoT or image data like building plans, should be performed. The evaluation of the different approaches should be based on a dataset with a much broader variety of views and which could be made available as a benchmark dataset for upcoming research to consolidate the results of this work.

Second, the developed image XAI method, Grad-Ram, wheras it is sensitive toward model and data, lacks fidelity. While the field of CV XAI methods for regression tasks is still emerging with limited methods published, also new methods should be developed, overcoming the limitations of Grad-Ram. This is necessary to effectively use visual XAI methods for CNNs performing continuous regression tasks. Furthermore, the methods should also be adaptable to MVNNs. Moreover, also additional XAI methods should be developed that can be used independently for a large variety of data sources, like the combination of tabular and image data to explain MVNNs better, answering the call of Barredo Arrieta et al. (2020) for more explainability methods in this area. With better explainability, the usage of black box models for high stakes decision could be enabled and users would gain more trust in ML models.

Third, this thesis used mainly metrical evaluations for models and XAI methods. While metrical evaluations measure the quality objectively, practical evaluations should be incorporated in the future, e.g., conducting a whole HCXAI process in real estate appraisal (e.g., online platform or AVM software vendor). By this, user groups, their opinion, and requirements can be better evaluated and incorporated compared to a metrical evaluation process. This may lead to additional insights into the user groups and a better understanding of the socio-technical system.

Lastly, additional replication and ablation studies are necessary. While the datasets used mainly focus on the U.S. (Philadelphia and Asheville), because of their data availability, additional tests should be conducted if multi-view real estate appraisal is beneficial also in other regions, e.g., Europe or Asia. Furthermore, the role of GIS data should be inspected more to clarify precisely which GIS data are helpful for real estate appraisal.

6.2. Outlook

Multi-view learning can bring multiple benefits to real estate appraisal. AVMs have improved predictive power by using previously omitted factors relating to soft information, e.g., style and aesthetics. From an economic view, multi-view learning and the use of GIS and images as data sources help transform Experience qualities of goods into Search qualities, reducing information asymmetries and strengthening the market in the light of the SEC Theory. Despite these benefits, an essential drawback of multi-view AVMs is their intransparency. Therefore, we developed different modeling strategies and XAI methods to derive an explanation of the price from the AVM. As reported in related work, multi-view learning will become even more important in real estate appraisal. Potentially, it will be used beyond real estate appraisal. We hope to inspire practitioners and researchers to apply multi-view learning to other pricing applications for used goods like cars or financial products like stocks.

Part II.

Publications

7. A Comparison of Multi-View Learning Strategies for Satellite Image-based Real Estate Appraisal

Jan-Peter Kucklick, Oliver Müller

– Paderborn University (UPB)

Authors	Kucklick, J-P. & Müller, O.
Year	2021
Type	Conference Paper
Outlet	The AAAI-21 Workshop on Knowledge Discovery from Unstructured Data in Financial Services
Rating (VHB Jourqual 3)	-
Contribution	Jan-Peter Kucklick (90%) & Oliver Müller (10%)
Citation	Kucklick, J.-P., & Müller, O. (2021). A comparison of multi-view learning strategies for satellite image-based real estate appraisal. In *The AAAI-21 Workshop on Knowledge Discovery from Unstructured Data in Financial Services.* Retrieved from https://aaai-kdf.github.io/kdf2021/assets/pdfs/KDF_21_paper_12.pdf

8. Quantifying the Impact of Location Data for Real Estate Appraisal - A GIS-Based Deep Learning Approach

Jan-Peter Kucklick, Jennifer Müller, Daniel Beverungen, Oliver Müller

– Paderborn University (UPB)

Authors	Kucklick, J-P., Müller, J., Beverungen, D. & Müller, O.
Year	2021
Type	Conference Paper
Outlet	European Conference on Information Systems (ECIS)
Rating (VHB Jourqual 3)	B
Contribution	Jan-Peter Kucklick (45%), Jennifer Müller (35%), Daniel Beverungen (10%) & Oliver Müller (10%)
Citation	Kucklick, J.-P., Müller, J., Beverungen, D., & Müller, O. (2021). Quantifying the impact of location data in real estate appraisal - a gis-based deep learning approach. In *Proceedings of the twenty-first European Conference on Information Systems (ECIS 2021), virtual, 14-16 June.* Retrieved from https://aisel.aisnet.org/ecis2021_rip/23/

9. Visual Interpretability of Image-based Real Estate Appraisal

Jan-Peter Kucklick

– Paderborn University (UPB)

Authors	Kucklick, J-P.
Year	2022
Type	Conference Paper
Outlet	Hawaii International Conference on System Sciences
Rating (VHB Jourqual 3)	C
Contribution	Jan-Peter Kucklick (100%)
Citation	Kucklick, J.-P. (2022b). Visual Interpretability of Image-based Real Estate Appraisal. In *Proceedings of the 55th Hawaii International Conference on System Science (HICSS-55), virtual, 4-7 January* (pp. 1510-1519). doi: 10.24251/HICSS.2022.187

10. Tackling the Accuracy–Interpretability Trade-off: Interpretable Deep Learning Models for Satellite Image-based Real Estate Appraisal

Jan-Peter Kucklick, Oliver Müller

– Paderborn University (UPB)

Authors	Kucklick, J-P. & Müller, O.
Year	2023
Type	Journal Paper
Outlet	ACM Transactions on Management Information Systems (TMIS)
Rating (VHB Jourqual 3)	B
Contribution	Jan-Peter Kucklick (80%) & Oliver Müller (20%)
Citation	Kucklick, J.-P., & Müller, O. (2023). Tackling the accuracy-interpretability trade-off: Interpretable deep learning models for satellite image-based real estate appraisal. *ACM Transactions on Management Information Systems*, 14(1). doi: 10.1145/3567430

11. Elucidating the Predictive Power of Search and Experience Qualities for Pricing of Complex Goods – A Machine Learning-based Study on Real Estate Appraisal

Jan-Peter Kucklick, Jennifer Priefer, Daniel Beverungen, Oliver Müller

– Paderborn University (UPB)

Authors	Kucklick, J-P., Priefer, J., Beverungen, D., Müller, O.
Year	Under review
Type	Journal Paper
Outlet	Information Systems Frontiers
Rating (VHB Jourqual 3)	B
Contribution	Jan-Peter Kucklick (45%), Jennifer Priefer (40%), Daniel Beverungen (10%), Oliver Müller (5%)
Citation	Kucklick, J.-P., Priefer, J., Beverungen, D., & Müller, O. (under review). Elucidating the Predictive Power of Search and Experience Qualities for Pricing of Complex Goods – A Machine Learning-based Study on Real Estate Appraisal. *Information Systems Frontiers.*

12. HIEF: A Holistic Interpretability and Explainability Framework

Jan-Peter Kucklick

– Paderborn University (UPB)

Authors	Kucklick, J-P.
Year	2023
Type	Journal Paper
Outlet	Journal of Decision Systems (JDS)
Rating (VHB Jourqual 3)	B
Contribution	Jan-Peter Kucklick (100%)
Citation	Kucklick, J.-P. (2023). Hief: A holistic interpretability and explainability framework. *Journal of Decision Systems*, 1-41. doi: 10.1080/12460125.2023.2207268
Comment	A previous version was published as: Kucklick, J.-P. (2022). Towards a model- and data-focused taxonomy of XAI systems. In *17th International Conference on Wirtschaftsinformatik (WI-22), virtual, 21-23 February* (pp. 1– 7). Retrieved from https://aisel.aisnet.org/wi2022/business_analytics/business_analytics/2

References

Abidoye, R. B., & Chan, A. P. C. (2018). Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal*, *24*(1), 71–83. doi: 10.1080/14445921.2018.1436306

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. doi: 10.1109/ACCESS.2018.2870052

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). Sanity checks for saliency maps. In *Advances in Neural Information Processing Systems* (Vol. 31). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2018/file/294a8ed24b1ad22ec2e7efea049b8737-Paper.pdf

Ahmed, E., & Moustafa, M. (2016). House price estimation from visual and textual features. *8th International Conference on Neural Computation Theory and Applications*. Retrieved from https://arxiv.org/pdf/1609.08399.pdf?amp=1 doi: 10.5220/0006040700620068

Akerlof, G. A. (1978). The market for “lemons”: Quality uncertainty and the market mechanism. In P. Diamond & M. Rothschild (Eds.), *Uncertainty in Economics* (p. 235-251). Academic Press. Retrieved from https://www.sciencedirect.com/science/article/pii/B978012214850750022X doi: https://doi.org/10.1016/B978-0-12-214850-7.50022-X

Albina, A. R. (2019). Assessing the impact of a GIS for improving novice crisis decision-making. In *Proceedings of the Twenty-fifth Americas Conference on Information Systems (AMCIS 2019), Cancun, 15-17 August.* Retrieved from https://aisel.aisnet.org/amcis2019/data_science_analytics_for_decision_support/data_science_analytics_for_decision_support/13

Alrwais, O., Hilton, B., Horan, T., & Bechor, T. (2016). An organizational perspective on gis payoffs for the public sector: Is usage the missing link? In *Proceedings of the Twenty-second Americas Conference on Information Systems (AMCIS 2016), San Diego, 11-14 August.* Retrieved from https://aisel.aisnet.org/amcis2016/

`Decision/Presentations/9`

Alvarez-Melis, D., & Jaakkola, T. S. (2018). Towards robust interpretability with self-explaining neural networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (p. 7786–7795). Red Hook, NY, USA: Curran Associates Inc.

Andrews, R. L., & Srinivasan, T. C. (1995). Studying consideration effects in empirical choice models using scanner panel data. *Journal of Marketing Research*, *32*(1), 30–41.

Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and explanation in ai-informed decision making. *Machine Learning and Knowledge Extraction*, *4*(2), 556–579. Retrieved from `https://www.mdpi.com/2504-4990/4/2/26` doi: 10.3390/make4020026

Anselin, L. (1998). Gis research infrastructure for spatial analysis of real estate markets. *Journal of Housing Research*, *9*(1), 113–133. doi: 10.1080/10835547.1998.12091930

Arslan, H., & Farkas, D. (2016). Using gis to understand the relationship of community factors and police shootings in the united states: a first look. In *Proceedings of the Twenty-second Americas Conference on Information Systems (AMCIS 2016), San Diego, 11-14 August.* Retrieved from `https://aisel.aisnet.org/amcis2016/ITProj/Presentations/25`

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, *22*(2), 325–352. doi: 10.17705/1jais.00664

Au, Q., Herbinger, J., Stachl, C., Bischl, B., & Casalicchio, G. (2022). Grouped feature importance and combined features effect plot. *Data Mining and Knowledge Discovery*, *36*(4), 1401–1450. Retrieved from `https://doi.org/10.1007/s10618-022-00840-5` doi: 10.1007/s10618-022-00840-5

Augasta, M. G., & Kathirvalavakumar, T. (2012). Reverse engineering the neural networks for rule extraction in classification problems. *Neural Processing Letters*, *35*(2), 131–150. Retrieved from `https://doi.org/10.1007/s11063-011-9207-8` doi: 10.1007/s11063-011-9207-8

Azizi, I., & Rudnytskyi, I. (2022). Improving real estate rental estimations with visual data. *Big Data and Cognitive Computing*, *6*(3). Retrieved from `https://www.mdpi.com/2504-2289/6/3/96` doi: 10.3390/bdcc6030096

Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied sciences*, *8*(11), 2321. doi: 10.3390/app8112321

Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(2), 423-443. doi: 10.1109/TPAMI.2018.2798607

Bansal, N., Agarwal, C., & Nguyen, A. (2020). Sam: The sensitivity of attribution methods to hyperparameters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8673–8683).

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... Herrera, F. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, *58*, 82–115. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1566253519308103` doi: https://doi.org/10.1016/j.inffus.2019.12.012

Bayer, S., Gimpel, H., & Markgraf, M. (2021). The role of domain expertise in trusting and following explainable AI decision support systems. *Journal of Decision Systems*, 1–29. Retrieved from `https://doi.org/10.1080/12460125.2021.1958505` doi: 10.1080/12460125.2021.1958505

Belle, V., & Papantonis, I. (2021). Principles and practice of explainable machine learning. *Frontiers in Big Data*, *4*. Retrieved from `https://www.frontiersin.org/articles/10.3389/fdata.2021.688969` doi: 10.3389/fdata.2021.688969

Bency, A. J., Rallapalli, S., Ganti, R. K., Srivatsa, M., & Manjunath, B. S. (2017). Beyond spatial auto-regressive models: Predicting housing prices with satellite imagery. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 320–329). doi: 10.1109/WACV.2017.42

Berente, N., Seidel, S., & Safadi, H. (2019). Research commentary—data-driven computationally intensive theory development. *Information Systems Research*, *30*(1), 50-64. Retrieved from `https://doi.org/10.1287/isre.2018.0774` doi: 10.1287/isre.2018.0774

Bessinger, Z., & Jacobs, N. (2016). Quantifying curb appeal. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 4388–4392). doi: 10.1109/ICIP.2016.7533189

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013, nov). Digital business strategy: Toward a next generation of insights. *MIS Q*, *37*(2), 471–482. Retrieved from `http://www.jstor.org/stable/43825919`

Biao, Z., Gaodi, X., Bin, X., & Canqiang, Z. (2012). The effects of public green spaces on residential property value in Beijing. *Journal of Resources and Ecology*, *3*(3), 243–252.

Bin, J., Gardiner, B., Li, E., & Liu, Z. (2020). Multi-source urban data fusion for property value assessment: A case study in Philadelphia. *Neurocomputing*, *404*, 70–83. doi: j.neucom.2020.05.013

Bin, J., Gardiner, B., Liu, Z., & Li, E. (2019). Attention-based multi-modal fusion for improved real estate appraisal: a case study in Los Angeles. *Multimedia Tools and Applications*, *78*(22), 31163–31184. doi: 10.1007/s11042-019-07895-5

Bin, Y., & Kumbier, K. (2020). Veridical Data Science. *Proceedings of the National Academy of Sciences*, *117*(8), 3920–3929. doi: 10.1073/pnas.1901326117

Bostrom, R. P., & Heinen, J. S. (1977). Mis problems and failures: A socio-technical perspective, part ii: The application of socio-technical theory. *MIS Q*, *1*(4), 11–28. Retrieved 2023-03-06, from `http://www.jstor.org/stable/249019`

Bourassa, S., Cantoni, E., & Hoesli, M. (2010). Predicting house prices with spatial dependence: a comparison of alternative methods. *Journal of Real Estate Research*, *32*(2), 139–160.

Breath, C. M., & Ives, B. (1986). Competitive information systems in support of pricing. *MIS Q*, *10*(1), 85–96. doi: 10.2307/248883

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324

Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-8*(6), 679–698. doi: 10.1109/TPAMI.1986.4767851

Caron, M., & Müller, O. (2020). Hardening soft information: A transformer-based approach to forecasting stock return volatility. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 4383–4391). doi: 10.1109/BigData50022.2020.9378134

Castellano, G., Kessous, L., & Caridakis, G. (2008). Emotion recognition through multiple modalities: Face, body gesture, speech. In C. Peter & R. Beale (Eds.), *Affect and emotion in human-computer interaction: From theory to applications* (pp. 92–103). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from `https://doi.org/10.1007/978-3-540-85099-1_8` doi: 10.1007/978-3-540-85099-1_8

Chancellor, S. (2023). Toward Practices for Human-Centered Machine Learning. *Commun. ACM*, *66*(3), 78–85. Retrieved from `https://doi.org/10.1145/3530987` doi: 10.1145/3530987

Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 839–847).

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Q*, *36*(4), 1165–1188. Retrieved from `http://www.jstor.org/stable/41703503` doi: 10.2307/41703503

Chen, J., Evans-Cowley, J. S., Rutherford, R. C., & Stanley, B. W. (2013). An empirical analysis of effect of housing curb appeal on sales price of newer houses. *Applied Finance*, *1407*.

Chen, L., Wu, Y., DSouza, A. M., Abidin, A. Z., Wismüller, A., & Xu, C. (2018). MRI tumor segmentation with densely connected 3D CNN. In E. D. Angelini & B. A. Landman (Eds.), *Medical Imaging 2018: Image Processing* (Vol. 10574, p. 105741F). SPIE. Retrieved from `https://doi.org/10.1117/12.2293394` doi: 10.1117/12.2293394

Chiles, B. (2021). Shrouded prices and firm reputation: Evidence from the u.s. hotel industry. *Management Science*, *67*(2), 964–983. doi: 10.1287/mnsc.2019.3566

Chocarro, R., Cortinas, M., & Villanueva, M. L. (2021). Different channels for different services: information sources for services with search, experience and credence attributes. *The Service Industries Journal*, *41*(3-4), 261–284. doi: 10.1080/02642069.2018.1508457

Chollet, F. (2017). *Deep learning with python.* New York, NY: Manning Publications.

Chrisman, N. R. (1999). What does 'GIS'mean? *Transactions in GIS*, *3*(2), 175–186.

City of Asheville. (2018). *Real estate appraisal residential building 2018.* `http://data-avl.opendata.arcgis.com/datasets/bunco::real-estate-appraisal-residential-building-2018`.

City of Philadelphia. (2018). *Metadata cataloug.* `https://metadata.phila.gov/`.

Clark, P. (2021, nov). *Zillow Shuts Home-Flipping Business After Racking Up Losses.* Retrieved from `https://www.bloomberg.com/news/articles/2021-11-02/zillow-shuts-down-home-flipping-business-after-racking-up-losses`

Cohen, J. P., & Coughlin, C. C. (2008). Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of Regional Science*, *48*(5), 859–878. doi: 10.1111/j.1467-9787.2008.00569.x

Comyns, B., & Figge, F. (2015). Greenhouse gas reporting quality in the oil and gas industry: A longitudinal study using the typology of "search","experience" and "credence" information. *Accounting, Auditing & Accountability Journal*, *28*(3), 403-433. doi: 10.1108/AAAJ-10-2013-1498

D'Acci, L. (2014). Monetary, subjective and quantitative approaches to assess urban quality of life and pleasantness in cities (hedonic price, willingness-to-pay, positional value, life satisfaction, isobenefit lines). *Social Indicators Research*, *115*(2), 531–559. doi: 10.1007/s11205-012-0221-7

D'Acci, L. (2019). Quality of urban area, distance from city centre, and housing value. case study on real estate values in turin. *Cities*, *91*, 71–92. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0264275118308552` doi: 10.1016/j.cities.2018.11.008

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, p. 886-893 vol. 1). doi: 10.1109/CVPR.2005.177

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of Law and Economics*, *16*(1), 67-88. Retrieved from `https://doi.org/10.1086/466756` doi: 10.1086/466756

Davenport, T. H., & Patil, D. J. (2012). Data scientist. *Harvard Business Review*, *90*(5), 70–76.

de Castro Neto, M., Nascimento, M., Sarmento, P., Ribeiro, S., Rodrigues, T., & Painho, M. (2018). Implementation of a dashboard for security forces data visualization. In *18.a Conferência da Associação Portuguesa de Sistemas de Informação, Santarém, 12-13 October.*

De Nadai, M., & Lepri, B. (2018). The economic value of neighborhoods: predicting real estate prices from the urban environment. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 323–330). doi: 10.1109/DSAA.2018.00043

Dimoka, Hong, & Pavlou. (2012). On product uncertainty in online markets: Theory and evidence. *MIS Q*, *36*(2), 395–426. doi: 10.2307/41703461

Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, *30*(4), 171–198. doi: 10.1257/jep.30.4.171

Donovan, G. H., & Butry, D. T. (2011). The effect of urban trees on the rental price of single-family homes in Portland, Oregon. *Urban Forestry & Urban Greening*, *10*(3), 163–168. doi: 10.1016/j.ufug.2011.05.007

Du, M., Liu, N., & Hu, X. (2019, dec). Techniques for interpretable machine learning. *Commun. ACM*, *63*(1), 68–77. Retrieved from `https://doi.org/10.1145/3359786` doi: 10.1145/3359786

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 196–212). Cham: Springer International Publishing.

Dueker, K. J., & Kjerne, D. (1989). *Multipurpose cadastre: Terms and definitions.* American Soc. for Photogrammetry and Remote Sensing.

Dulleck, U., Kerschbamer, R., & Sutter, M. (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *American Economic Review*, *101*(2), 526–555. doi: 10.1257/aer.101.2.526

Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable ai: Towards a reflective sociotechnical approach. In C. Stephanidis, M. Kurosu, H. Degen, & L. Reinerman-Jones (Eds.), *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence* (pp. 449–466). Cham: Springer International Publishing.

Elam, E., & Stigarll, A. (2012). Landscape and house appearance impacts on the price of single-family houses. *Journal of Environmental Horticulture*, *30*(4), 182–188. doi: 10.24266/0738-2898.30.4.182

Elnagar, S., & Thomas, M. A. (2019). Real estate image-based appraisal using mask region based convolutional networks. In *Twenty-fifth Americas Conference on Information Systems (AMCIS 2019), Cancun, 15-17 August.* Retrieved from `https://aisel.aisnet.org/amcis2019/ai_semantic_for_intelligent_info_systems/ai_semantic_for_intelligent_info_systems/3`

Farkas, D., Hilton, B., Pick, J., Ramakrishna, H., Sarkar, A., & Shin, N. (2016). A tutorial on geographic information systems: A ten-year update. *Communications of the Association for Information Systems*, *38*(1), 190–234. doi: 10.17705/1CAIS.03809

Fathalla, A., Salah, A., Li, K., Li, K., & Francesco, P. (2020). Deep end-to-end learning for price prediction of second-hand items. *Knowledge and Information Systems*, *62*(12), 4541–4568. Retrieved from `https://doi.org/10.1007/s10115-020-01495-8` doi: 10.1007/s10115-020-01495-8

Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair AI. *Business & Information Systems Engineering*, *62*(4), 379–384. Retrieved from `https://doi.org/10.1007/s12599-020-00650-3` doi: 10.1007/s12599-020-00650-3

Ford, G. T., Smith, D. B., & Swasy, J. L. (1988). An empirical test of the search, experience and credence attributes framework. *ACR North American Advances*, *15*, 239–244.

Förster, M., Hühn, P., Klier, M., & Kluge, K. (2022). User-centric explainable AI: design and evaluation of an approach to generate coherent counterfactual explanations for structured data. *Journal of Decision Systems*, 1–32. Retrieved from `https://doi.org/10.1080/12460125.2022.2119707` doi: 10.1080/12460125.2022.2119707

Förster, M., Klier, M., Kluge, K., & Sigler, I. (2020). Fostering human agency: a process for the design of user-centric XAI systems. In *Proceedings of the fortieth International Conference on Information Systems (ICIS 2020), virtual, 13-16 December.* Retrieved from `https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/12`

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0925527314004253` doi: https://doi.org/10.1016/j.ijpe.2014.12.031

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. Retrieved from `http://www.jstor.org/stable/2699986`

Fu, X., Jia, T., Zhang, X., Li, S., & Zhang, Y. (2019). Do street-level scene perceptions affect housing prices in Chinese megacities? An analysis using open access datasets and deep learning. *PloS one*, *14*(5), e0217505. doi: 10.1371/journal.pone.0217505

Gavrikov, P. (2020). *Visualkeras.* `https://github.com/paulgavrikov/visualkeras`.

George, R. (2020). *White and brown house.* `https://unsplash.com/de/fotos/9gGvNWBeOq4`.

Ghose. (2009). Internet exchanges for used goods: An empirical analysis of trade patterns and adverse selection. *MIS Q*, *33*(2), 263–291. doi: 10.2307/20650292

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. doi: 10.1080/10618600.2014.907095

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

Google LLC. (2022). *Get a static map.* Retrieved from `https://developers.google.com/maps/documentation/maps-static/start`

Gosiewska, A., Kozak, A., & Biecek, P. (2021). Simpler is better: Lifting interpretability-performance trade-off via automated feature engineering. *Deci-*

sion Support Systems, *150*. Retrieved from https://www.sciencedirect.com/science/article/pii/S016792362100066X doi: https://doi.org/10.1016/j.dss.2021.113556

Gregor, S. (2006). The nature of theory in information systems. *MIS Q*, *30*(3), 611–642. Retrieved from http://www.jstor.org/stable/25148742 doi: 10.2307/25148742

Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Q*, *23*(4), 497–530. Retrieved from http://www.jstor.org/stable/249487

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Q*, *37*(2), 337–355. Retrieved from http://www.jstor.org/stable/43825912

Gröbel, S., & Thomschke, L. (2018). Hedonic pricing and the spatial structure of housing data–an application to Berlin. *Journal of Property Research*, *35*(3), 185–208. Retrieved from https://doi.org/10.1080/09599916.2018.1510428 doi: 10.1080/09599916.2018.1510428

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, *51*(5), 1–42. Retrieved from https://doi.org/10.1145/3236009 doi: 10.1145/3236009

Guliker, E., Folmer, E., & van Sinderen, M. (2022). Spatial determinants of real estate appraisals in the netherlands: A machine learning approach. *ISPRS International Journal of Geo-Information*, *11*(2). Retrieved from https://www.mdpi.com/2220-9964/11/2/125 doi: 10.3390/ijgi11020125

Gunasti, K., Kara, S., & Ross Jr, W. T. (2020). Effects of search, experience and credence attributes versus suggestive brand names on product evaluations. *European Journal of Marketing*, *54*(12), 2941–2963. doi: doi.org/10.1108/EJM-12-2017-0972

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. Retrieved from https://doi.org/10.1126/scirobotics.aay7120 doi: 10.1126/scirobotics.aay7120

Han, B. (2022, jun). *Real estate industry accounted for 16.9% of GDP in 2021.* Retrieved from https://www.realtrends.com/articles/real-estate-industry-accounted-for-16-9-of-gdp-in-2021/

Han, L., Yin, Z., Xia, Z., Guo, L., Tang, M., & Jin, R. (2019). Vision-based price suggestion for online second-hand items. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 1988–1996). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3343031.3350936

Hao, L., & Fan, M. (2014). An analysis of pricing models in the electronic book market. *MIS Q*, *38*(4), 1017–1032. doi: 10.25300/MISQ/2014/38.4.04

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction* (2nd ed.). Springer. Retrieved from `http://www-stat.stanford.edu/$\sim$tibs/ElemStatLearn/`

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 1026–1034).

He, K., Zhang, X., Ren, S., & Sun, J. (2016a). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778).

He, K., Zhang, X., Ren, S., & Sun, J. (2016b). Identity mappings in deep residual networks. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 630–645). Cham: Springer International Publishing.

Helber, P., Bischke, B., Guo, Q., Hees, J., & Dengel, A. (2019). Multi-scale machine learning for the classification of building property values. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4873–4876). doi: 10.1109/IGARSS.2019.8900257

Helbich, M., Jochem, A., Mücke, W., & Höfle, B. (2013). Boosting the predictive accuracy of urban hedonic house price models through airborne laser scanning. *Computers, Environment and Urban Systems*, *39*, 81–92. doi: 10.1016/j.compenvurbsys.2013.01.001

Herm, L.-V., Heinrich, K., Wanner, J., & Janiesch, C. (2023). Stop ordering machine learning algorithms by their explainability! a user-centered investigation of performance and explainability. *International Journal of Information Management*, *69*, 102538. Retrieved from `https://www.sciencedirect.com/science/article/pii/S026840122200072X` doi: https://doi.org/10.1016/j.ijinfomgt.2022.102538

Herm, L.-V., Wanner, J., Seubert, F., & Janiesch, C. (2021). I don't get it, but it seems valid! The connection between explainability and comprehensibility in (X) AI research. In *Proceedings of the twenty-first European Conference on Information Systems (ECIS 2021), virtual, 14-16 June.* Retrieved from `https://aisel.aisnet.org/ecis2021_rp/82/`

Heuwinkel, T., Kucklick, J.-P., & Müller, O. (2022). Using geolocated text to quantify location in real estate appraisal. In *Proceedings of the 55th Hawaii International Conference on System Science (HICSS-55), virtual, 4-7 January* (pp. 5744–5753). Retrieved from `http://hdl.handle.net/10125/80039` doi: 10.24251/HICSS.2022.700

Hill, R. J., & Scholz, M. (2018). Can geospatial data improve house price indexes? a hedonic imputation approach with splines. *Review of Income and Wealth*, *64* (4), 737–756. doi: 10.1111/roiw.12303

Hinz, Hann, & Spann. (2011). Price discrimination in e-commerce? an examination of dynamic pricing in name-your-own price markets. *MIS Q*, *35* (1), 81–98. doi: 10.2307/23043490

Ho, J. (2016). Machine learning for causal inference: An application to air quality impacts on house prices. *arXiv: 1808.02547 v1*.

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., ... Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595* (7866), 181–188. Retrieved from `https://doi.org/10.1038/s41586-021-03659-0` doi: 10.1038/s41586-021-03659-0

Iacobucci, D. (1992). An empirical examination of some basic tenets in services: goods-services continua. *Advances in Services Marketing and Management*, *1*, 23–52.

Jaakonmäki, R., Müller, O., & Vom Brocke, J. (2017). The impact of content, context, and creator on user engagement in social media marketing. In *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS-50), Hawaii, 4-7 January.* Retrieved from `http://hdl.handle.net/10125/41289` doi: 10.24251/HICSS.2017.136

Jameel, M., Arif, M. u. I., Hintsches, A., & Schmidt-Thieme, L. (2021). Automation of leasing vehicle return assessment using deep learning models. In Y. Dong, D. Mladenić, & C. Saunders (Eds.), *Machine Learning and Knowledge Discovery in Databases: Applied Data Science Track* (pp. 259–274). Cham: Springer International Publishing. doi: 10.1007/978-3-030-67667-4_16

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York, NY: Springer.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353* (6301), 790–794. doi: 10.1126/science.aaf7894

Jha, S. B., Pandey, V., Jha, R. K., & Babiceanu, R. F. (2020). Machine learning approaches to real estate market prediction problem: A case study. *arXiv preprint arXiv:2008.09922*.

Joblove, G. H., & Greenberg, D. (1978). Color spaces for computer graphics. In *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques* (p. 20–25). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/800248.807362` doi: 10.1145/800248.807362

Johnson, E. B., Tidwell, A., & Villupuram, S. V. (2020). Valuing curb appeal. *The Journal of Real Estate Finance and Economics*, *60*(1), 111–133. doi: 10.1007/s11146-019-09713-z

Kaminski, M. E. (2019). The right to explanation, explained. *Berkeley Tech. LJ*, *34*, 189-217. doi: 10.15779/Z38TD9N83H

Kamm, M. R., Kucklick, J.-P., Schneider, J., & vom Brocke, J. (2021). Data mining for small shops: Empowering brick-and-mortar stores through BI functionalities of a loyalty program1. *Information Systems Management*, *38*(4), 270–286. Retrieved from `https://doi.org/10.1080/10580530.2020.1855486` doi: 10.1080/10580530.2020.1855486

Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2020). Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*, 104919. doi: j.landusepol.2020.104919

Keenan, P., Carroll, P., Conway, H., & Lee, S. (2017). Spatial modelling of climate change in Irish forestry. In *Proceedings of the Twenty-third Americas Conference on Information Systems (AMCIS 2017), Boston, 10-12 August.*

Kim, J., & Kim, K.-m. (2022). How the pattern recognition ability of deep learning enhances housing price estimation. *Journal of the Economic Geographical Society of Korea*, *25*, 183–201. doi: 10.23841/egsk.2022.25.1.183

Kiranshastry. (n.d.). *Haus.* `https://www.flaticon.com/de/kostenloses-icon/haus_845022?term=haus&page=1&position=1&origin=search&related_id=845022.`

Kisilevich, S., Keim, D., & Rokach, L. (2013). A gis-based decision support system for hotel room rate estimation and temporal price prediction: The hotel brokers' context. *Decision Support Systems*, *54*(2), 1119–1133. doi: 10.1016/j.dss.2012.10.038

Klein, L. R. (1998). Evaluating the potential of interactive media through a new lens: Search versus experience goods. *Journal of Business Research*, *41*(3), 195-203. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0148296397000623` doi: https://doi.org/10.1016/S0148-2963(97)00062-3

Koch, D., Despotovic, M., Leiber, S., Sakeena, M., Döller, M., & Zeppelzauer, M.

(2019). Real estate image analysis: a literature review. *Journal of Real Estate Literature*, *27*(2), 269–300. doi: 10.22300/0927-7544.27.2.269

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. doi: 10.3905/jpm.2017.43.6.202

Kostic, Z., & Jevremovic, A. (2020). What image features boost housing market predictions? *IEEE Transactions on Multimedia*, *22*(7), 1904–1916. doi: 10.1109/TMM.2020.2966890

Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*, *281*(3), 628–641. Retrieved from https://www.sciencedirect.com/science/article/pii/S0377221719307581 doi: https://doi.org/10.1016/j.ejor.2019.09.018

Kucklick, J.-P. (2022a). Towards a model- and data-focused taxonomy of XAI systems. In *17th International Conference on Wirtschaftsinformatik (WI-22), virtual, 21-23 February* (pp. 1–7). Retrieved from https://aisel.aisnet.org/wi2022/business_analytics/business_analytics/2

Kucklick, J.-P. (2022b). Visual interpretability of image-based real estate appraisal. In *Proceedings of the 55th Hawaii International Conference on System Science (HICSS-55), virtual, 4-7 January* (pp. 1510–1519). Retrieved from http://hdl.handle.net/10125/79519 doi: 10.24251/HICSS.2022.187

Kucklick, J.-P. (2023). Hief: a holistic interpretability and explainability framework. *Journal of Decision Systems*, 1-41. Retrieved from https://doi.org/10.1080/12460125.2023.2207268 doi: 10.1080/12460125.2023.2207268

Kucklick, J.-P., Kamm, M. R., Schneider, J., & Vom Brocke, J. (2020). Extending loyalty programs with bi functionalities. In *Proceedings of the 53th Hawaii International Conference on System Science (HICSS-53), Hawaii, 7-10 January* (pp. 168–177). Retrieved from http://hdl.handle.net/10125/63761

Kucklick, J.-P., Müller, J., Beverungen, D., & Müller, O. (2021). Quantifying the impact of location data in real estate appraisal - a gis-based deep learning approach. In *Proceedings of the twenty-first European Conference on Information Systems (ECIS 2021), virtual, 14-16 June.* Retrieved from https://aisel.aisnet.org/ecis2021_rip/23/

Kucklick, J.-P., & Müller, O. (2020). Location, location, location: Satellite image-based real-estate appraisal. *16th Symposium on Statistical Challenges in Electronic Commerce Research.* Retrieved from https://arxiv.org/abs/2006.11406

Kucklick, J.-P., & Müller, O. (2023). Tackling the accuracy-interpretability trade-off: Interpretable deep learning models for satellite image-based real estate appraisal. *ACM Trans. Manage. Inf. Syst.*, *14*(1). Retrieved from `https://doi.org/10.1145/3567430` doi: 10.1145/3567430

Kucklick, J.-P., & Müller, O. (2021). A comparison of multi-view learning strategies for satellite image-based real estate appraisal. In *The AAAI-21 Workshop on Knowledge Discovery from Unstructured Data in Financial Services.* Retrieved from `https://aaai-kdf.github.io/kdf2021/assets/pdfs/KDF_21_paper_12.pdf`

Kucklick, J.-P., Priefer, J., Beverungen, D., & Müller, O. (under review). Elucidating the predictive power of search and experience qualities for pricing of complex goods – A machine learning-based study on real estate appraisal. *Information Systems Frontiers*.

Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., & Szopinski, D. (2022). An update for taxonomy designers. *Business & Information Systems Engineering*, *64*, 421–439. Retrieved from `https://doi.org/10.1007/s12599-021-00723-x` doi: 10.1007/s12599-021-00723-x

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132–157. doi: 10.1086/259131

Lansford Jr, N. H., & Jones, L. L. (1995). Marginal price of lake recreation and aesthetics: an hedonic approach. *Journal of Agricultural and Applied Economics*, *27*(1), 212–223. doi: 10.1017/S107407080001974X

Law, S., & Neira, M. (2019). An unsupervised approach to geographical knowledge discovery using street level and street network images. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery* (p. 56–65). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3356471.3365238

Law, S., Paige, B., & Russell, C. (2019). Take a look around: Using street view and satellite images to estimate house prices. *ACM Trans. Intell. Syst. Technol.*, *10*(5), 1–19. Retrieved from `https://doi.org/10.1145/3342240` doi: 10.1145/3342240

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., ... Alstyne, M. V. (2009). Computational social science. *Science*, *323*(5915), 721-723. Retrieved from `https://www.science.org/doi/abs/10.1126/science.1167742` doi: 10.1126/science.1167742

Leby, J. L., & Hashim, A. H. (2010). Liveability dimensions and attributes: Their relative importance in the eyes of neighbourhood residents. *Journal of Construction* *in Developing Countries*, *15*(1), 67–91.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. doi: 10.1038/nature14539

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551. doi: 10.1162/neco.1989.1.4.541

Lee, C., & Park, K.-H. (2021). Using photographs and metadata to estimate house prices in South Korea. *Data Technologies and Applications*, *55*(2), 280–292. Retrieved from `https://doi.org/10.1108/DTA-05-2020-0111` doi: 10.1108/DTA-05-2020-0111

Lee, S., & Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, *41*(6), 3041–3046. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0957417413008518` doi: https://doi.org/10.1016/j.eswa.2013.10.034

Leffrang, D., & Müller, O. (2021). Should i follow this model? the effect of uncertainty visualization on the acceptance of time series forecasts. In *IEEE VIS 2021 TREX Workshop* (pp. 1–7). IEEE. Retrieved from `https://trexvis.github.io/Workshop2021/papers/Leffrang.pdf`

Lestari, A., Rumantir, G., & Tapper, N. (2016). A spatio-temporal analysis on the forest fire occurrence in central Kalimantan, Indonesia. In *Proceedings of the twenties Pacific Asia Conference On Information Systems (PACIS 2016), Chiayi, 27 June-1 July.* Retrieved from `https://aisel.aisnet.org/pacis2016/90/`

Li, Y., Yang, M., & Zhang, Z. (2018). A survey of multi-view representation learning. *IEEE Transactions on Knowledge and Data Engineering*, *31*(10), 1863–1883. doi: 10.1109/TKDE.2018.2872063

Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790*.

Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and soft. *Review of Corporate Finance Studies*, *8*(1), 1–41. Retrieved from `https://doi.org/10.1093/rcfs/cfy009` doi: 10.1093/rcfs/cfy009

Ligus, M., & Peternek, P. (2016). Measuring structural, location and environmental effects: A hedonic analysis of housing market in wroclaw, poland. *Procedia-Social and Behavioral Sciences*, *220*, 251–260. doi: 10.1016/j.sbspro.2016.05.497

Limsombunchai, V. (2004). House price prediction: hedonic price model vs. artificial neural network. In *New zealand agricultural and resource economics society conference* (pp. 25–26).

Lin, R. F.-Y., Ou, C., Tseng, K.-K., Bowen, D., Yung, K. L., & Ip, W. H. (2021). The Spatial neural network model with disruptive technology for property appraisal in real estate industry. *Technological Forecasting and Social Change*, *173*, 121067. Retrieved from https://www.sciencedirect.com/science/article/pii/S0040162521004996 doi: https://doi.org/10.1016/j.techfore.2021.121067

Lipton, Z. C. (2018). The Mythos of Model Interpretability. *Queue*, *16*(3), 31–57. doi: 10.1145/3236386.3241340

Liu, X., Xu, Q., Yang, J., Thalman, J., Yan, S., & Luo, J. (2018). Learning multi-instance deep ranking and regression network for visual house appraisal. *IEEE Transactions on Knowledge and Data Engineering*, *30*(8), 1496–1506. doi: 10.1109/TKDE.2018.2791611

Lorenz, F., Willwersch, J., Cajias, M., & Fuerst, F. (2022). Interpretable machine learning for real estate market analysis. *Real Estate Economics*. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-6229.12397 doi: https://doi.org/10.1111/1540-6229.12397

Lukyanenko, R., Castellanos, A., Samuel, B., Tremblay, M., & Maass, W. (2021). Research Agenda for Basic Explainable AI. In *Proceedings of the Twenty-seventh Americas Conference on Information Systems (AMCIS 2021), virtual, 9-13 August.* Retrieved from https://aisel.aisnet.org/amcis2021/art_intel_sem_tech_intelligent_systems/art_intel_sem_tech_intelligent_systems/7/

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 4765–4774). Red Hook, NY, USA: Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., ... others (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, *2*, 749–760. doi: 10.1038/s41551-018-0304-0

Martens, D., & Provost, F. (2014). Explaining data-driven document classifications. *MIS Q*, *38*(1), 73–100. Retrieved from https://www.jstor.org/stable/26554869 doi: 10.25300/MISQ/2014/38.1.04

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60–68.

McCluskey, W., Deddis, W., Mannis, A., McBurney, D., & Borst, R. (1997). Interactive application of computer assisted mass appraisal and geographic information systems. *Journal of Property Valuation and Investment*, *15*(5), 448–465. doi: doi.org/10.1108/14635789710189227

McCluskey, W. J., Daud, D. Z., & Kamarudin, N. (2014). Boosted regression trees: An application for the mass appraisal of residential property in malaysia. *Journal of Financial Management of Property and Construction*, *19*(2), 152–167.

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. doi: 10.1080/09599916.2013.781204

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2020). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management*, *39*(1), 53–63. Retrieved from `https://doi.org/10.1080/10580530.2020.1849465` doi: 10.1080/10580530.2020.1849465

Microsoft Inc. (2020). *Get a static map.* "https://docs.microsoft.com/en-us/bingmaps/rest-services/imagery/get-a-static-map".

Mitra, K., Reiss, M. C., & Capella, L. M. (1999). An examination of perceived risk, information search and behavioral intentions in search, experience and credence services. *Journal of Services Marketing*, *13*(2), 208–228. doi: 10.1108/08876049910273763

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable ai systems. *ACM Trans. Interact. Intell. Syst.*, *11*(3–4), 1–45. Retrieved from `https://doi.org/10.1145/3387166` doi: 10.1145/3387166

Molnar, C. (2019). *Interpretable machine learning.* Research Triangle, NC: Lulu Press, Inc.

Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K.-R. (2019). Layer-wise relevance propagation: An overview. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable ai: Interpreting, explaining and visualizing deep learning* (pp. 193–209). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-28954-6_10` doi: 10.1007/978-3-030-28954-6_10

Müller, O., Junglas, I., vom Brocke, J., & Debortoli, S. (2016). Utilizing big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information Systems*, *25*(4), 289–302. Retrieved from `https://doi.org/10.1057/ejis.2016.2` doi: 10.1057/ejis.2016.2

Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, *263*(2), 611–624. Retrieved from https://www.sciencedirect.com/science/article/pii/S0377221717304332 doi: https://doi.org/10.1016/j.ejor.2017.05.005

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore - predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (pp. 779–785).

Naser, N., Serte, S., & Al-Turjman, F. (2021). From traditional house price appraisal to computer vision-based: A survey. In E. Ever & F. Al-Turjman (Eds.), *Forthcoming networks and sustainability in the iot era* (pp. 1–10). Cham: Springer International Publishing.

National Association of Home Builders. (2020). *Housing's contribution to gross domestic product.* National Association of Home Builders (NAHB). Retrieved from https://www.nahb.org/news-and-economics/housing-economics/Housings-Economic-Impact/Housings-Contribution-to-Gross-Domestic-Product

National Association of Realtors. (2021). *Real estate in a digital age* (Tech. Rep.). National Association of REALTORS. Retrieved from https://www.nar.realtor/sites/default/files/documents/2021-real-estate-in-a-digital-age-10-05-2021.pdf

National Association of Realtors. (2022). *2022 Appraisal Survey* (Tech. Rep.). Retrieved from https://cdn.nar.realtor//sites/default/files/documents/2022-appraisal-survey-09-27-2022.pdf

Naumzik, C., & Feuerriegel, S. (2020). One picture is worth a thousand words? the pricing power of images in e-commerce. In *Proceedings of The Web Conference 2020* (p. 3119–3125). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3366423.3380086 doi: 10.1145/3366423.3380086

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, *78*(2), 311-329. Retrieved from https://doi.org/10.1086/259630 doi: 10.1086/259630

Ng, A. (2021). *A chat with Andrew on MLOps: From model-centric AI.* DeepLearningAI. Retrieved from https://www.youtube.com/watch?v=06-AZXmwHjo

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, *22*(3), 336–359. Retrieved from https://doi.org/10.1057/

`ejis.2012.26` doi: 10.1057/ejis.2012.26

Noor, N. M., Asmawi, M. Z., & Abdullah, A. (2015). Sustainable urban regeneration: Gis and hedonic pricing method in determining the value of green space in housing area. *Procedia-Social and Behavioral Sciences*, *170*, 669–679. Retrieved from `https://doi.org/10.1016/j.sbspro.2015.01.069` doi: 10.1016/j.sbspro.2015.01.069

Noor, N. M., & Rosni, N. A. (2013). Determination of spatial factors in measuring urban sprawl in Kuantan using remote sensing and GIS. *Procedia-Social and Behavioral Sciences*, *85*, 502–512. Retrieved from `https://doi.org/10.1016/j.sbspro.2013.08.379` doi: 10.1016/j.sbspro.2013.08.379

Nouriani, A., & Lemke, L. (2022). Vision-based housing price estimation using interior, exterior & satellite images. *Intelligent Systems with Applications*, 200081. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2667305322000217` doi: https://doi.org/10.1016/j.iswa.2022.200081

Nussberger, A.-M., Luo, L., Celis, L. E., & Crockett, M. J. (2022). Public attitudes value interpretability but prioritize accuracy in Artificial Intelligence. *Nature Communications*, *13*(1), 5821. Retrieved from `https://doi.org/10.1038/s41467-022-33417-3` doi: 10.1038/s41467-022-33417-3

Oberdorf, F., Schaschek, M., Stein, N., & Flath, C. (2021). Neural Process Mining: Multi-Headed Predictive Process Analytics in Practice. In *Proceedings of the twenty-first European Conference on Information Systems (ECIS 2021), virtual, 14-16 June.* Retrieved from `https://aisel.aisnet.org/ecis2021_rp/54/`

OECD. (2021). *Tax revenues of subsectors of general government as % of total tax revenue.* `https://stats.oecd.org/Index.aspx?DataSetCode=Rev#`. (Accessed on 26.02.2021)

Oh, & Lucas. (2006). Information technology and pricing decisions: Price adjustments in online computer markets. *MIS Q*, *30*(3), 755–775. doi: 10.2307/25148748

Onnom, W., Tripathi, N., Nitivattananon, V., & Ninsawat, S. (2018). Development of a liveable city index (LCI) using multi criteria geospatial modelling for medium class cities in developing countries. *Sustainability*, *10*(2), 520–539. Retrieved from `https://doi.org/10.3390/su10020520` doi: 10.3390/su10020520

OpenDataPhilly. (2011). *Your source for open data in the Philadelphia region.* Retrieved from `https://www.opendataphilly.org`

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003, jan). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, *21*(4), 383–401. Retrieved from `https://doi.org/10.1108/`

14635780310483656 doi: 10.1108/14635780310483656

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert Systems with Applications*, *42*(6), 2928–2934. doi: 10.1016/j.eswa.2014.11.040

Park, S., Kim, S., & Lee, J. (2020). A pilot study on the relationship between urban green spaces and fine particulate matter. *International Journal of Geospatial and Environmental Research*, *7*(1), 3. Retrieved from https://dc.uwm.edu/ijger/vol7/iss1/3

Park, T., Kim, M., Jang, C., Choung, T., Sim, K.-A., Seo, D., & Chang, S. I. (2018). The public health impact of road-traffic noise in a highly-populated city, Republic of Korea: Annoyance and Sleep Disturbance. *Sustainability*, *10*(8), 2947. Retrieved from https://doi.org/10.3390/su10082947 doi: 10.3390/su10082947

Passchier-Vermeer, W., & Passchier, W. F. (2000). Noise exposure and public health. *Environmental health perspectives*, *108*(suppl 1), 123–131. Retrieved from https://doi.org/10.1289/ehp.00108s1123 doi: 10.1289/ehp.00108s1123

Peña, S., Fuentes, C. M., Cervera, L. E., & Hernández, V. (2012). Planning support systems: a computer-assisted mass appraisal (CAMA) system for Ciudad Juárez, México. *Journal of Property Tax Assessment & Administration*, *9*(4), 25–40. Retrieved from https://researchexchange.iaao.org/jptaa/vol9/iss4/2

Peterson, K. (1993). Spatial decision support systems for real estate investment analysis. *International Journal of Geographical Information Science*, *7*(4), 379–392. Retrieved from https://doi.org/10.1080/02693799308901968 doi: 10.1080/02693799308901968

Peterson, S., & Flanagan, A. (2009). Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research*, *31*(2), 147–164. Retrieved from https://doi.org/10.1080/10835547.2009.12091245 doi: 10.1080/10835547.2009.12091245

Phan, T. D. (2018). Housing price prediction using machine learning algorithms: The case of melbourne city, australia. In *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)* (pp. 35–42). doi: 10.1109/iCMLDE.2018.00017

Phillips, R., Şimşek, A. S., & van Ryzin, G. (2015). The effectiveness of field price discretion: Empirical evidence from auto lending. *Management Science*, *61*(8), 1741–1759. doi: 10.1287/mnsc.2014.2084

Potrawa, T., & Tetereva, A. (2022). How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. *Journal* *of Business Research*, *144*, 50–65. Retrieved from `https://www.sciencedirect.com/science/article/pii/S014829632200039X` doi: https://doi.org/10.1016/j.jbusres.2022.01.027

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W., & Wallach, H. (2021). Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–52). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3411764.3445315` doi: 10.1145/3411764.3445315

Poursaeed, O., Matera, T., & Belongie, S. (2018). Vision-based real estate price estimation. *Machine Vision and Applications*, *29*(4), 667–676. doi: 10.1007/s00138-018-0922-2

Priefer, J. (2023). Geographic information systems in information systems research-review and research prospects. In *Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS-56), Hawaii, 3-6 January* (pp. 5048–5058). Retrieved from `https://hdl.handle.net/10125/103252`

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: unbiased boosting with categorical features. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 31). Curran Associates, Inc. Retrieved from `https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf`

Qu, G., Xiao, L., Hu, W., Wang, J., Zhang, K., Calhoun, V. D., & Wang, Y.-P. (2021). Ensemble manifold regularized multi-modal graph convolutional network for cognitive ability prediction. *IEEE Transactions on Biomedical Engineering*, *68*(12), 3564-3573. doi: 10.1109/TBME.2021.3077875

Quiring, S. M. (1987). Package your home to sell. *Leaflet/Texas Agricultural Extension Service; no. 2256.*.

Recker, J. (2021). *Scientific research in information systems: A beginner's guide* (2021st ed.). Berlin, Germany: Springer. doi: https://doi.org/10.1007/978-3-642-30048-6

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why S should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2939672.2939778` doi: 10.1145/

2939672.2939778

Rosemann, M., & Vessey, I. (2008). Toward improving the relevance of information systems research to practice: The role of applicability checks. *MIS Q*, *32*(1), 1–22. Retrieved from http://www.jstor.org/stable/25148826 doi: 10.2307/25148826

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. doi: 10.1086/260169

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *The American Economic Review*, *63*(2), 134–139. Retrieved 2023-02-23, from http://www.jstor.org/stable/1817064

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. doi: https://doi.org/10.1038/s42256-019-0048-x

Sah, V., Conroy, S. J., & Narwold, A. (2016). Estimating school proximity effects on housing prices: The importance of robust spatial controls in hedonic estimations. *The Journal of Real Estate Finance and Economics*, *53*(1), 50–76.

Samek, W., Binder, A., Montavon, G., Lapuschkin, S., & Müller, K.-R. (2016). Evaluating the visualization of what a deep neural network has learned. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(11), 2660–2673. doi: 10.1109/TNNLS.2016.2599820

Samek, W., & Müller, K.-R. (2019). Towards explainable artificial intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable ai: Interpreting, explaining and visualizing deep learning* (pp. 5–22). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-28954-6_1 doi: 10.1007/978-3-030-28954-6_1

Sarip, A. G., Hafez, M. B., & Daud, M. N. (2016). Application of fuzzy regression model for real estate price prediction. *Malaysian Journal of Computer Science*, *29*(1), 15–27. doi: 10.22452/mjcs.vol29no1.2

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3). Retrieved from https://doi.org/10.1007/s42979-021-00592-x doi: 10.1007/s42979-021-00592-x

Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems*, *29*(4), 260–278. Retrieved from https://doi.org/10.1080/12460125.2020.1819094 doi: 10.1080/12460125.2020.1819094

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017, oct). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626).

Sheehan, E., Meng, C., Tan, M., Uzkent, B., Jean, N., Burke, M., ... Ermon, S. (2019). Predicting economic development using geolocated wikipedia articles. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2698–2706). Retrieved from https://doi.org/10.1145/3292500.3330784 doi: 10.1145/3292500.3330784

Shimomura, M., Nakaune, M., Yamada, T., Tsukazawa, K., & Nakamura, K. (2020). Crop yields prediction with cnn and visual explanation with grad-ram. In *IEICE Conferences Archives.*

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289 – 310. Retrieved from https://doi.org/10.1214/10-STS330 doi: 10.1214/10-STS330

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Q*, *35*(3), 553–572. Retrieved from http://www.jstor.org/stable/23042796 doi: 10.2307/23042796

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sing, T. F., Yang, J. J., & Yu, S. M. (2021). Boosted tree ensembles for artificial intelligence based automated valuation models (ai-avm). *The Journal of Real Estate Finance and Economics*, *65*, 649–674. Retrieved from https://doi.org/10.1007/s11146-021-09861-1 doi: 10.1007/s11146-021-09861-1

Smith, S. (2013, 08 16). *Morning headlines: Cnn money says philly's best neighborhood is mt. airy.* Retrieved from https://www.phillymag.com/property/2013/08/16/looking-for-phillys-best-neighborhood-get-your-mt-airy-groove-on/

Soh, Markus, & Goh. (2006). Electronic marketplaces and price transparency: Strategy, information technology, and success. *MIS Q*, *30*(3), 705–723. doi: 10.2307/25148746

Solovev, K., & Pröllochs, N. (2021). Integrating floor plans into hedonic models for rent price appraisal. In *Proceedings of the Web Conference 2021* (p. 2838–2847). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3442381.3449967 doi: 10.1145/3442381.3449967

Steiger, D. M. (1998). Enhancing user understanding in a decision support system: A theoretical basis and framework. *Journal of Management Information*

Systems, *15*(2), 199-220. Retrieved from `https://doi.org/10.1080/07421222.1998.11518214` doi: 10.1080/07421222.1998.11518214

Stupak, J. M. (2019). *Introduction to u.s. economy: Housing market.* https://fas.org/sgp/crs/misc/IF11327.pdf. Congressional Research Service (CRS).

Sugumaran, V., & Mobley, L. R. (2000). Integrating spatial regression into decision support systems. *Proceedings of the sixth American Conference on Information Systems (AMCIS 2000), Long Beach, 10-13 August*, 249. Retrieved from `https://aisel.aisnet.org/amcis2000/249/`

Suksa-Ngiam, W., Mbugua, J., & Chatterjee, S. (2016). A GIS decision support system for crop cultivation. In *Proceedings of the Twenty-second Americas Conference on Information Systems (AMCIS 2016), San Diego, 11-14 August.* Retrieved from `https://aisel.aisnet.org/amcis2016/Decision/Presentations/5/`

Sun, L., & Zhu, H. (2009). GIS-Based spatial decision support system for real estate appraisal. In *2009 Fourth International Conference on Computer Sciences and Convergence Information Technology* (pp. 1339–1344).

Sun, S. (2013). A survey of multi-view machine learning. *Neural Computing and Applications*, *23*(7-8), 2031–2038. Retrieved from `https://doi.org/10.1007/s00521-013-1362-6` doi: 10.1007/s00521-013-1362-6

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (p. 2818-2826). Los Alamitos, CA, USA: IEEE Computer Society. Retrieved from `https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.308` doi: 10.1109/CVPR.2016.308

Theriault, M., Des Rosier, F., & Vandersmissen, M. (1999). *GIS-based simulation of accessibility to enhance hedonic modeling and property value appraisal: An application to Quebec City metropolitan area* [Working Papers]. Retrieved from `https://EconPapers.repec.org/RePEc:fth:lavadm:99-011`

Tomsett, R., Harborne, D., Chakraborty, S., Gurram, P., & Preece, A. (2020). Sanity checks for saliency metrics. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(04), 6021-6029. Retrieved from `https://ojs.aaai.org/index.php/AAAI/article/view/6064` doi: 10.1609/aaai.v34i04.6064

Tsagris, M., & Fafalios, S. (2022). Advanced car price modelling and prediction. In M. K. Terzioğlu (Ed.), *Advances in econometrics, operational research, data science and actuarial studies: Techniques and theories* (pp. 479–494). Cham:

Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-85254-2_29` doi: 10.1007/978-3-030-85254-2_29

van Heijst, D., Potharst, R., & van Wezel, M. (2008). A support system for predicting ebay end prices. *Decision Support Systems*, *44*(4), 970–982. doi: 10.1016/j.dss.2007.11.004

Vanags, J., Geipele, I., Sarkans, A., & Usenieks, D. (2017). Housing heterogeneity dimensions and their elements: a systemic approach. *Baltic Journal of Real Estate Economics and Construction Management*, *5*(1), 23–37. doi: 10.1515/bjreecm-2017-0003

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (p. 6000–6010). Red Hook, NY, USA: Curran Associates Inc.

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, *2*(1), 63-84. Retrieved from `https://doi.org/10.1287/isre.2.1.63` doi: 10.1287/isre.2.1.63

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144. Retrieved from `https://doi.org/10.1016/j.jsis.2019.01.003` doi: 10.1016/j.jsis.2019.01.003

Vujcic, M., Tomicevic-Dubljevic, J., Zivojinovic, I., & Toskovic, O. (2019). Connection between urban green areas and visitors' physical and mental well-being. *Urban forestry & urban greening*, *40*, 299–307. Retrieved from `https://doi.org/10.1016/j.ufug.2018.01.028` doi: 10.1016/j.ufug.2018.01.028

Wang, P.-Y., Chen, C.-T., Su, J.-W., Wang, T.-Y., & Huang, S.-H. (2021). Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. *IEEE Access*, *9*, 55244–55259. doi: 10.1109/ACCESS.2021.3071306

Wang, S., Zhu, J., Yin, Y., Wang, D., Cheng, T. C. E., & Wang, Y. (2021). Interpretable multi-modal stacking-based ensemble learning method for real estate appraisal. *IEEE Transactions on Multimedia*, 1. doi: 10.1109/TMM.2021.3126153

Wang, Z., & Yang, J. (2018). Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, 514–521.

Wanner, J., Herm, L.-V., Heinrich, K., & Janiesch, C. (2021). Stop ordering machine learning algorithms by their explainability! an empirical investigation of the tradeoff between performance and explainability. In D. Dennehy, A. Griva, N. Pouloudi, Y. K. Dwivedi, I. Pappas, & M. Mäntymäki (Eds.), *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society* (pp. 245–258). Cham: Springer International Publishing.

Wanner, J., Herm, L.-V., Heinrich, K., & Janiesch, C. (2022). A social evaluation of the perceived goodness of explainability in machine learning. *Journal of Business Analytics*, *5*(1), 29–50. Retrieved from https://doi.org/10.1080/2573234X.2021.1952913 doi: 10.1080/2573234X.2021.1952913

Wanner, J., Herm, L.-V., Heinrich, K., Janiesch, C., & Zschech, P. (2020). White, grey, black: Effects of XAI augmentation on the confidence in AI-based decision support systems. In *Proceedings of the fortieth International Conference on Information Systems (ICIS 2020), virtual, 13-16 December.* Retrieved from https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/14/

Wastensteiner, J., Weiss, T. M., Haag, F., & Hopf, K. (2021). Explainable ai for tailored electricity consumption feedback–an experimental evaluation of visualizations. In *Proceedings of the twenty-first European Conference on Information Systems (ECIS 2021), virtual, 14-16 June.* Retrieved from https://aisel.aisnet.org/ecis2021_rp/55/

Wei, C., Fu, M., Wang, L., Yang, H., Tang, F., & Xiong, Y. (2022). The research development of hedonic price model-based real estate appraisal in the era of big data. *Land*, *11*(3). Retrieved from https://www.mdpi.com/2073-445X/11/3/334 doi: 10.3390/land11030334

Wilson, J. Q., & Kelling, G. E. (1982). Broken windows: The police and neighborhood safety. *Atlantic Monthly*, *249*(3), 29–38.

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. Retrieved from https://doi.org/10.1080/21650020.2019.1704429 doi: 10.1080/21650020.2019.1704429

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3–19).

Wu, Z., Chen, R., Meadows, M. E., Sengupta, D., & Xu, D. (2019). Changing urban green spaces in Shanghai: trends, drivers and policy implications. *Land Use Policy*, *87*, 104080. Retrieved from https://doi.org/10.1016/j.landusepol.2019.104080 doi: 10.1016/j.landusepol.2019.104080

Xu, C., Tao, D., & Xu, C. (2013). A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*.

Xu, Z., Tao, D., Zhang, Y., Wu, J., & Tsoi, A. C. (2014). Architectural style classification using multinomial latent logistic regression. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 600–615). Cham: Springer International Publishing.

Yan, X., Hu, S., Mao, Y., Ye, Y., & Yu, H. (2021). Deep multi-view learning methods: A review. *Neurocomputing*, *448*, 106-129. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0925231221004768` doi: https://doi.org/10.1016/j.neucom.2021.03.090

Yang, R. R., Chen, S., & Chou, E. (2018). AI blue book: Vehicle price prediction using visual features. *arXiv preprint arXiv:1803.11227*.

Yiorkas, C., & Dimopoulos, T. (2017). Implementing GIS in real estate price prediction and mass valuation: the case study of Nicosia District. In *Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017)* (Vol. 10444, p. 104440F).

You, Q., Pang, R., Cao, L., & Luo, J. (2017). Image-based appraisal of real estate properties. *IEEE Transactions on Multimedia*, *19*(12), 2751–2759. doi: 10.1109/TMM.2017.2710804

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 818–833). Cham: Springer International Publishing.

Zhang, M. L., & Han, G. C. (2012). Application of spatial decision support system in real estate appraisal. In *Applied mechanics and materials* (Vol. 174, pp. 3286–3290). Retrieved from `https://doi.org/10.4028/www.scientific.net/AMM.174-177.3286` doi: 10.4028/www.scientific.net/AMM.174-177.3286

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2022). What makes a good image? Airbnb demand analytics leveraging interpretable image features. *Management Science*, *68*(8), 5644-5666. Retrieved from `https://doi.org/10.1287/mnsc.2021.4175` doi: 10.1287/mnsc.2021.4175

Zhao, J., Liu, X., Kuang, Y., Chen, Y. V., & Yang, B. (2018). Deep cnn-based methods to evaluate neighborhood-scale urban valuation through street scenes perception. In *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)* (pp. 20–27). doi: 10.1109/DSC.2018.00012

Zhao, J., Xie, X., Xu, X., & Sun, S. (2017). Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, *38*, 43–54. doi: 10.1016/

j.inffus.2017.02.007

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016, June). Learning deep features for discriminative localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (p. 2921-2929). Los Alamitos, CA, USA: IEEE Computer Society. Retrieved from https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.319 doi: 10.1109/CVPR.2016.319

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1452-1464. doi: 10.1109/TPAMI.2017.2723009

Zillow Group Inc. (2020). *What is a Zestimate? Zillow's Zestimate Accuracy.* https://www.zillow.com/z/zestimate/.