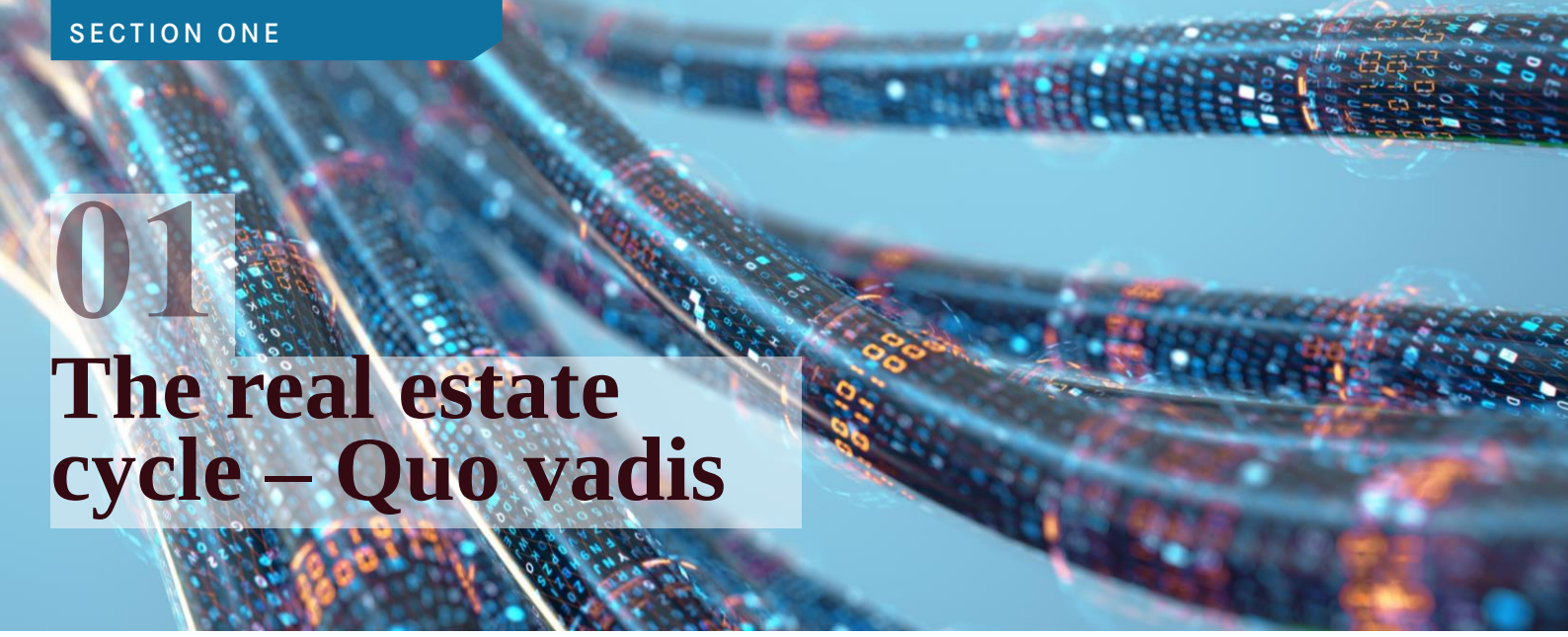
RESEARCH PAPER

Data Intelligence and Real Estate
Machines are the real game changer

Dr. Marcelo Cajias
Dr. Anett Wins

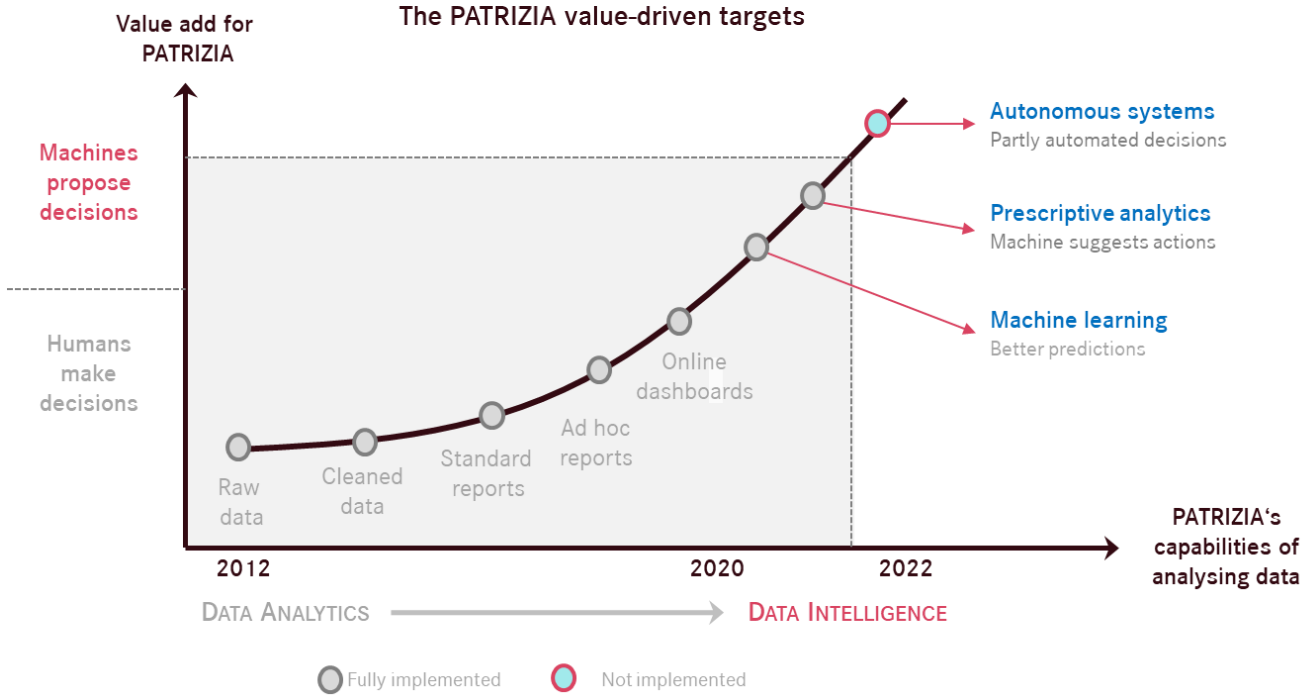
PATRIZIA



01 The real estate cycle – Quo vadis

The real estate industry is evolving faster than ever and has learned that its current strategic setup must adapt in the near future. The way of doing business with real estate will change, and investment managers' success might inversely relate to their speed in adopting new technologies to make real estate management more efficient, transparent and sustainable. However, the current market environment feels like a mix of actors with a clear mindset between "business as usual" and "it is now or never." In other words, new technologies are not being equally adapted by the industry. This is indeed a strong assumption and a disruptive thought at the same time. Still, real estate today requires the ability to deal with digitalisation, machine learning (ML), tokenisation, and IoT, to name some. These concepts are certainly in vogue and represent the real estate industry's increasing need to add a new pillar to the foundation: real estate data intelligence.

The past decade has shed light on one of the industry's darkest problems: lack of data. The real estate industry is more the rule than the exception regarding a lack of transparency. Perhaps, therefore, data collection has been the focus of many think tanks. One might think that the increase in information would lead to greater transparency, more accurate pricing, and, finally, more efficient markets. But the opposite seems to be the case as increased access to data has made understanding assets much more difficult. This is primarily because traditional performance indicators are now confronted with new and alternative data sources outside the real estate valuation books. Yet, in the era of google street view, open street maps, IoT, web scraping, and satellite pictures, the pricing and assessment of a property is more complicated than described by brokers' exposés and the last valuation report.

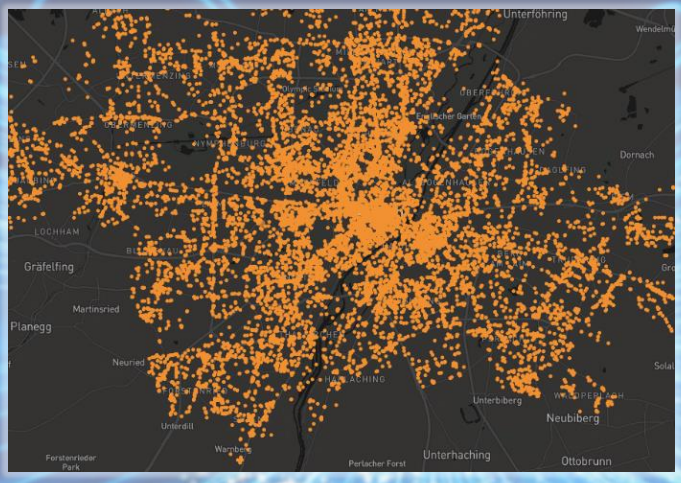


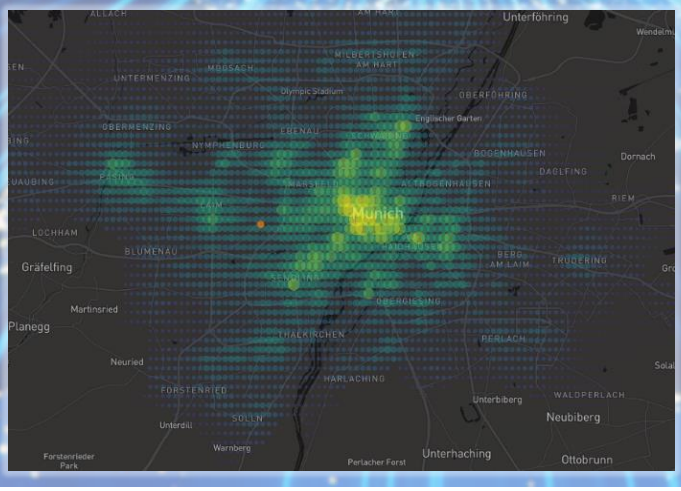
DIGITAL ASSESSMENT OF OPPORTUNITIES

Artificial intelligence can answer the question of the quality of an investment's location within a city. This usually involves collecting socio-economic data about the city on a postcode level and information regarding access to public transportation, educational institutions or restaurants. The assessment traditionally also considers factors such as experience and instinct. But while the amount of spatial detail has been growing exponentially since the introduction of google maps and open street maps, the intelligence needed to process all this information and use it to benefit real estate investments remains in its infancy.

What would happen if an intelligent machine could view spatial data through the eyes of a real estate investor? To make this happen, the first step is to teach the machine which types of amenities should be scored positively or negatively, for example, a park vs a nuclear power plant. In other words, the machine needs access to a dictionary of amenities with corresponding definitions in real estate investment terms.
The second step is to apply a decision rule for assessing the amenities in a market, that is, an algorithm. Once the machine screens all amenities in the market and identifies locations with an oversupply and undersupply of positive or negative amenities, the results need to be standardized from 0 to 100, with 100 indicating the best location.

The map of above shows all Munich's daily used points of interest. From this explorative point of view, traffic nodes and popular urban areas can be identified. The algorithm takes distance to and importance of amenities into account: For example, living near public transport and other amenities used daily (such as educational institutions, shops, cafés and restaurants) significantly enhances residential areas. The result of the scoring algorithm is visualised on the map at the bottom: The more yellow and the larger the grid points, the more attractive are nearby locations for residential purposes.





What we feel

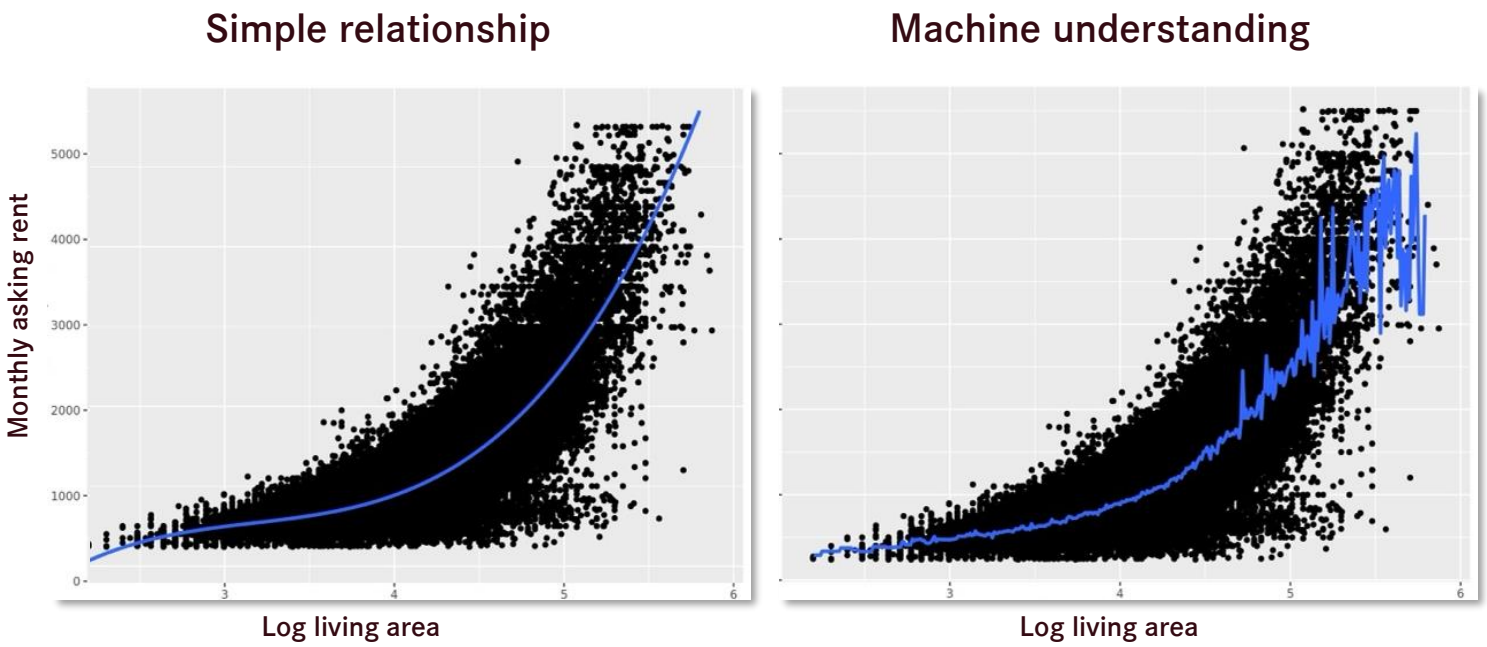
Nowadays, algorithms measure the attractiveness of locations within minutes, leading to cost savings, faster decision processes and a transparent and comprehensive assessment. This can make understanding real estate challenging as we need to trust the algorithm to validate the knowledge of local experts. The positive aspect of any algorithm is that it can confirm our feelings of a locations. More importantly, it can alert us to anomalies that we don't have on our radar. The primary usage, therefore, is not to replace specialists but to support them with knowledge.

What we don't see

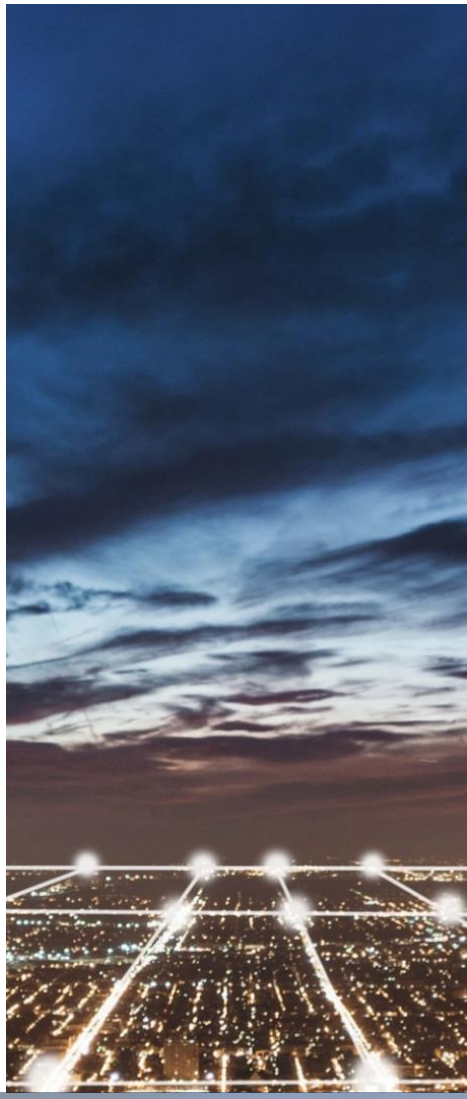
02 The breakthrough in understanding rents

The field of ML has developed rapidly over the last two decades. ML is an intelligent methodology for explaining the behaviour of a target variable with the help of computational power. While the simple regression tries to derive the elasticity of rents to changes in size via a single coefficient, the ML approach derives a complex non-linear relationship. The benefit of the non-linearity is the strong understanding of the complex relationships and the lower forecasting error compared to traditional methods. The main criticism of ML is the black-box pattern in decisions, that is, the lack of transparency in how a prediction is generated. As a response, the field of eXplainable Artificial Intelligence (XAI) has been developed and launched for the most important statistical languages after 2018. XAI targets methods for understanding the decision process in ML predictions to gain new insights into the importance of factors influencing the response variable, that is, rents.

Based on the asking rents of Munich, the linear regression approach (Y: rent; X: third-degree polynomial of logarithmized living area) establishes a polynomial relationship in which rents increase exponentially as the apartment size increases (cp. **Figure 2**, LHS). This looks obvious and intuitive. Even though the connection cannot be explained by economic theory, the ML approach (XGBoost) establishes a relationship with the same pattern but with higher accuracy and stronger variation (cp. **Figure 2**, RHS).



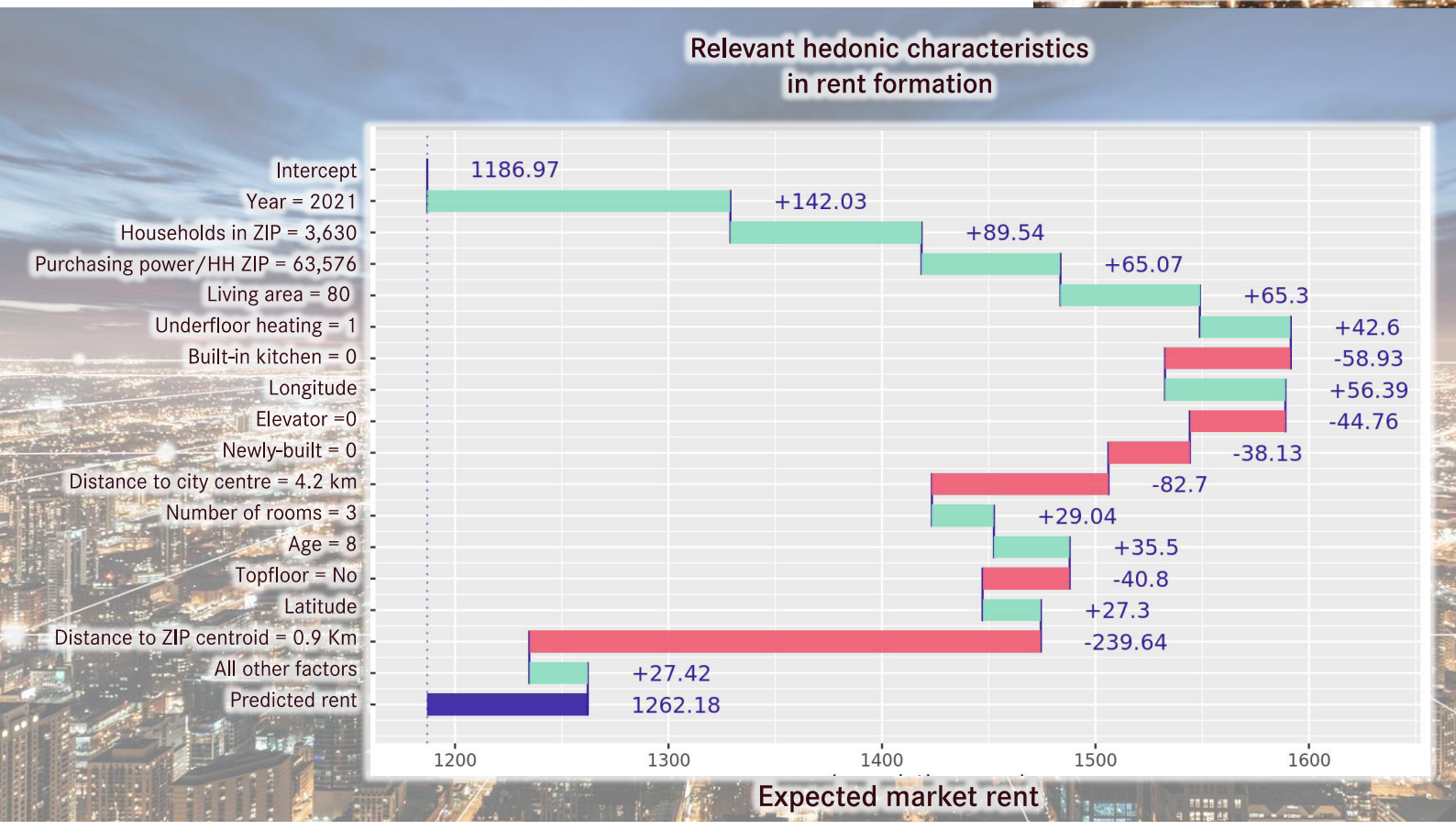
QUANTIFYING THE VALUE DRIVERS OF RENTS



The advanced ML model (multiple predictors) explains 88% of the rent variation, almost 15 percentage points more than the traditional approach. This extended knowledge makes benchmarking (new) portfolio assets more accurate and lessens the uncertainty of mispriced investment opportunities. From the perspective of an investment committee, an instrument that can reduce uncertainty is anything but “business as usual”.

Although ML models are accurate, model interpretation seems to be difficult at first glance. However, XAI sheds light on the complexity of rent prediction, which boosts the understanding of rent composition to a never seen scale. Any estimated rent can be decomposed into its single value drivers. For example, Figure 3 illustrates the break-down profile for an 80 sqm apartment with the following hedonic characteristics: underfloor heating, three rooms, and eight years old. The green and red bars indicate positive and negative contributions to the mean rent prediction, which are assigned to the explanatory characteristics. Based on an average rent of €1,186 (Intercept), the expected market rent for this apartment is €1,262 (blue bar). The underfloor heating (green) raises the expected rent but the lack of a built-in kitchen and elevator and the top floor (red) decrease the expected market rent value. Furthermore, time (that is, Year) and location specific characteristics (for example, population, wealth in ZIP and distances, etc.) affect the prediction.

What we now know



Centroid defined as the centre of the NUTS3 polygon

WHERE ARE WE TODAY?

Technology in all its facets is shaping the real estate industry. The change is not limited to single-use cases or controlled experiments but reaches the root of the business and, most importantly, how we understand real estate. The traditional focus on location[3], which evolved in the past decade to data[3] seems to be conquered by amenities[3] and artificial modeling[3]. This illustrates that it is now time to re-think the way we operate real estate. New methods and datasets entered the industry and were incorporated into investment decisions. Nowadays, we neither lack data nor knowledge, but we need to leave our comfort zone and learn from the outside to understand the real estate industry in a new way.

Dr. Marcelo Cajias

Head of Data Intelligence

Marcelo heads the Data Intelligence section at PATRIZIA. In his role he is responsible for the global portfolio of analytical solutions and dashboards that support strategic investment decisions by means of observed and unobserved machine learning forecast models for various asset classes. Marcelo studied business administration at the University of Regensburg in Germany, majoring in statistics, econometrics and real estate economics. He received his doctorate for his thesis on the economic impact of sustainability on listed real estate companies. As a dedicated researcher, his work has been published in various international journals and he was awarded the RICS Best Paper Award and the German Real Estate Research Prize.

Dr. Anett Wins

Data Scientist

Anett joined PATRIZIA in September 2020 as data scientist. She is responsible for building and estimating statistical models to find hidden patters and signals in data. She employs sophisticated machine learning algorithms to be used in predictive and prescriptive research questions. Anett studied statistics at the University of Munich and received her doctorate at the University of Augsburg for her thesis on the measurement of investors' preferences regarding sustainable and responsible investments.

A similar version of this article can be found in the Journal of Property Finance and Investment: https://doi.org/10.1108/JPIF-11-2021-0101