Leveraging data science in real estate

How data science can enable better decision making

a.s.r. real estate

July 2023

Leveraging data science in real estate

Searching for a needle in a haystack. Most people agree that using data science effectively can drive business success. At a.s.r. real estate we use data science to enhance return forecasting. With more accurate forecasts, we can make better investment decisions. However, growing data amounts and the complex data dependencies, make it challenging to reap the full benefits of data science. This paper examines both the added value and challenges of data science in real estate research, the conditions for success, and use cases.

Data science enables better decision-making

Data science can be used to translate large amounts of complex data into clear and actionable information. It is an interdisciplinary field that extracts insights from data using statistical analysis, machine learning and data visualisation.

This can lead to faster and better decision-making. For example, data science models may reveal opportunities for reducing maintenance costs or enhancing portfolio allocation and returns. Moreover, the accuracy of the data model is likely to improve with each year or quarter as more data is collected.

It is an attractive prospect, but it doesn't come for free or without risks. Implementing data science models can be challenging both from a technical and an organisational point of view. First of all, you need to integrate the data science toolkit into existing business processes. This integration requires careful consideration of how data science aligns with and enhances the overall organisational structure and workflows.

The accuracy of the data model is likely to improve with each year or quarter as more data is collected

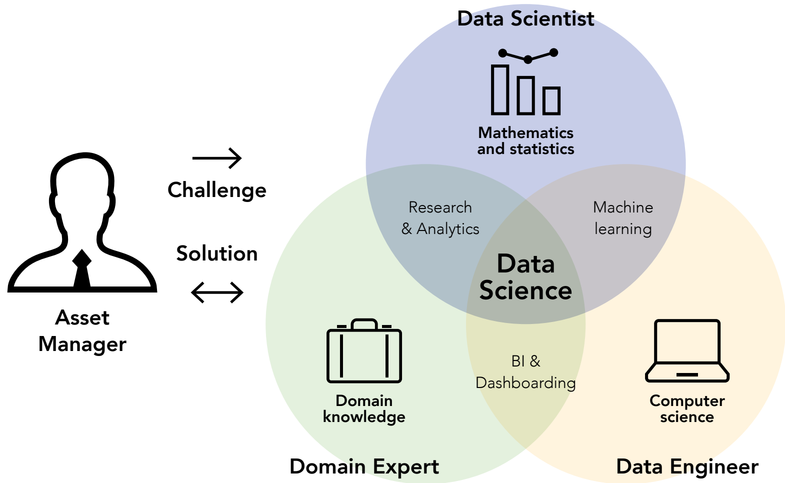
Collaboration between data scientists, domain experts and asset managers essential

Below are several considerations for the successful implementation of data science into existing business cycles. Additional considerations may apply. For example, the use of consumer data requires specific attention to the General Data Protection Regulation (GDPR) and data governance.

Team effort: Implementing data science initiatives is a collaborative task that involves four key roles, each with their own distinct responsibilities: the asset manager, the data scientist, the domain expert and the data engineer. Together, they engage in repeated and iterative conversations. The domain expert works closely with the asset manager to identify and understand the challenges at hand. The data scientist and domain expert explore data-driven possibilities for addressing these challenges. The data scientist takes charge of implementing analytical models and algorithms that extract insights from the data. They explore the data, identify patterns and develop predictive models. The data engineer is responsible for the collection and preparation of the data, as well as managing the infrastructure and implementing data integration processes. Building up a shared 'language' and understanding requires patience. Depending on the team's skills and capacity and the importance of the model in decision-making, it may be worth involving a data governance specialist, test engineer and data architect.

Data infrastructure: The integration of the model into the existing data infrastructure is crucial for optimising its use. This covers the entire process of uploading raw data to databases through the continuous updating of the deliverable (e.g., a real-time dashboard).

Figure 1 Roles and skills data science implementation



Model selection: Data science models come in many forms: supervised, unsupervised, and reinforcement learning. The desired level of transparency is a crucial factor in selecting the right model. Reinforcement learning models like neural networks are typically less transparent than their counterparts, making the models more challenging to interpret and comprehend. Other models are more transparent and easier to explain. Some models require large amounts of data, which tends to be challenging in real estate. Stakeholders should have a clear understanding of a model's capabilities and limitations and select the model type that best aligns with their shared goals.

In short, several organisational aspects are to be considered. Next up are the more technical considerations for choosing and implementing a data science model.

Data science models can improve forecasting

Expert opinion and the statistical analyses of market data (e.g. regression analysis) help us to construct our forecasts. We now use data science models to enhance these forecasts. We have a policy in place to ensure that data science models are used exclusively to support the forecasting of returns and are not used as autonomous forecasting tools.

Residential forecasting tool

The goal of our residential forecasting tool is to forecast the capital growth of the Dutch residential market over the next year. The model entails many macroeconomic, financial and real estate-related indicators (see examples in table 1). It uses a combination of free and paid data that is automatically accessible via APIs (application that allows automated data exchange). Most indicators have a lagged effect on capital growth ranging up to more than 1 year.

We chose a Random Forest classification model because it is both accurate and transparent. It consists of multiple decision trees and uses a voting mechanism to produce more accurate and robust results.

In the model, we separate market rental value growth and yield development. Both contribute to capital growth deviations, but they have different drivers. Forecasting gets truly interesting when it reveals differences in the occupier and investment markets cycle phase.

Table 1 Indicator sample residential forecasting tool

Category	Macroeconomic	Financial	Real estate
Indicator	Consumer trust	Coupon Dutch 10-yrs bond	Mortgage borrowing capacity

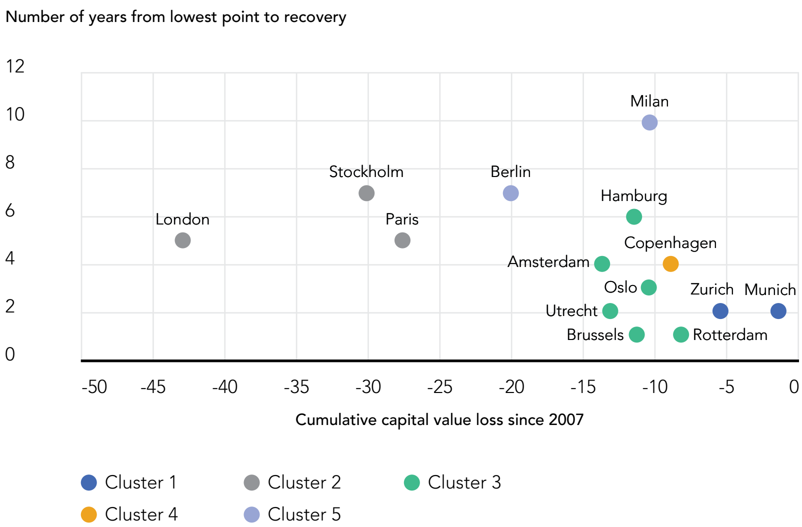
European retail market resiliency

In our European rental growth forecasting model, we found substantial differences in accuracy between sectors. Just before COVID-19, we proved that you can classify and cluster sector-market combinations by their crises resilience.

Some European retail markets take a big hit during a crisis but recover quickly (e.g., London). Others record a more limited devaluation of real estate but take longer to recover (e.g., Milan) (see figure 2). We considered three variables in this analysis:the number of years from the lowest point to recovery, the cumulative capital value loss since 2007 (both in figure) and the number of years from the beginning of the crisis to the lowest point. With these three variables we identified 5 different clusters. Institutional investors looking for diversification should spread their portfolio among these clusters.

We found that data science models produce the most accurate results in less policy-driven and more macroeconomic conditions

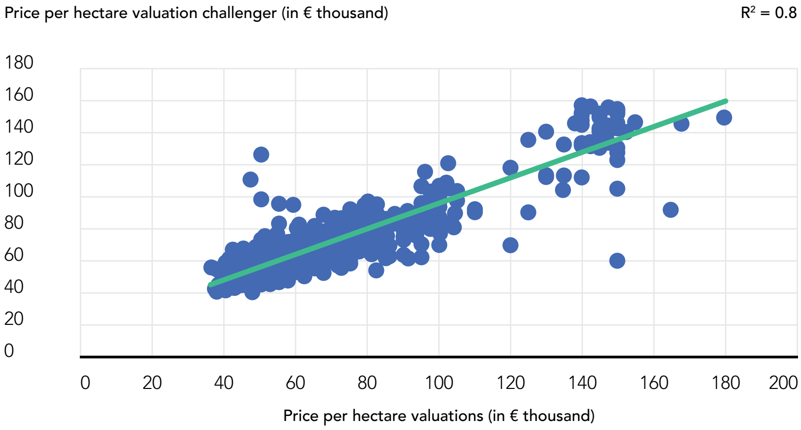
Figure 2 Cluster analysis of resiliency of European retail markets



Rural valuation challenger

The valuations of agricultural land are less transparent than those of traditional real estate markets. To comprehend the valuations of rural real estate and strategically plan acquisitions and dispositions, we need to understand what drives agricultural transaction prices. That is why we developed the valuation challenger, a gradient boosting model used to uncover relationships in the data. This machine learning algorithm combines weaker decision trees into a more accurate prediction model. The model can explain around 80% of transaction prices (see figure 3).

Figure 3 Accuracy rural valuation challenger



We found that locational characteristics (e.g., X and Y coordinates) are a better predictor of transaction prices than physical characteristics (e.g., type of land or % of peatland). Recent transaction prices of nearby agricultural land are also very important.

Context matters

We found that data science models produce the most accurate results in less policy-driven and more macroeconomic conditions. This means that you have to weigh the model forecast against the present market conditions and political context. At the time of writing, the Dutch residential market is impacted by proposed national policy, which is moving some investors to sell residential assets.In the current market conditions, other forecasting methods (e.g., comparative or scenario analyses) are more likely to produce reliable results than data science forecasting models (if used in isolation), as these models can't incorporate proposed policy shifts.

In conclusion, data science can yield very useful insights. But it is not the only helpful tool for forecasting real estate returns. Use data science forecasting models in combination with other forecasting methods and be aware of the macroeconomic and political context at the time of forecasting.

Awareness is key to unlocking the full potential of data science

The future of data science remains bright. Datafication will enhance existing models and will enable building new models. As with every new technique, new applications will improve decision-making. The many uses of AI-powered language models (e.g., ChatGPT) are yet to be fully uncovered and will make work more efficient for many businesses. Preventive maintenance of real estate will be possible once enough data is collected and analysed. Metropolitan and asset-level forecasting will improve as data becomes more accessible. The continued development of automatic valuation models will make most valuation processes quicker.

The true value of data science is only unlocked, however, when data science is integrated into business processes and all stakeholders are aware of the possibilities and limitations of the data science models in question.

Contact

For more information on the research vision, please contact:

Vinoo Khandekar
head of Research & Intelligence

T: +31 (0)6 23 90 13 75
E: vinoo.khandekar@asr.nl

a.s.r.
de nederlandse
verzekerings
maatschappij
voor alle
verzekeringen

www.asrrealestate.nl