**AI-based Real Estate Valuation: A Data-Driven Performance Analysis of Property and Neighborhood Influences**

Turtogtokh Shagai (https://orcid.org/0009-0004-4656-6867)

Dr. László Pitlik (<https://orcid.org/0000-0001-5819-0319>),

László Pitlik (Jr.) (<https://orcid.org/0000-0002-8058-9577>)

Mátyás Pitlik (<https://orcid.org/0000-0002-1991-3008>),

E-Mails: ssun41268@gmail.com, pitlik@my-x.hu, ptlklszl@my-x.hu, ppk@my-x.hu

Kodolányi János University and MY-X research team Hungary

# **Introduction**

The real estate market, inherently dynamic and multifaceted, has long relied on traditional methods of property valuation that often involve manual appraisals and heuristic approaches. These traditional methods, while established and widely used, are prone to biases and inconsistencies due to their subjective nature. The advent of Artificial Intelligence (AI) offers a transformative solution to these challenges by enabling the development of Automated Valuation Models (AVMs) that leverage large datasets and sophisticated analytics. Traditional real estate valuation methods have often been criticized for their reliance on manual processes, which introduce a significant degree of subjectivity and potential bias. These methods can be time-consuming and may not consistently reflect current market conditions. In contrast, AI technologies can process vast amounts of data quickly and objectively, leading to more accurate and consistent property valuations. The proposed COCO AI-based valuation model uses various property and neighborhood attributes to predict initial property prices. Significant property attributes such as size and the number of rooms are primary factors influencing predictions. Market fluctuations and unique house features are also incorporated to provide a comprehensive valuation. Data is collected from multiple sources, including market platforms, websites, and APIs, ensuring a robust dataset for analysis. The integration of property features, time series data, neighborhood dynamics, and market trends results in a more accurate and human error-free valuation model. By analyzing all relevant attributes — both property-specific and neighborhood-related — the model aims to achieve unbiased and accurate price predictions. This automated approach improves upon traditional manual methods, offering a cost-effective solution that benefits field experts and individuals making real estate investments. The ultimate goal is to automate the entire evaluation process, transitioning from manual-driven test cases to a fully automated system. This development will culminate in a bachelor's degree thesis, with the software becoming available as a web application for broader access.

# **Literature**

1. Real estate valuation

Real estate valuation is influenced by various factors, including property attributes, market trends, and participant perceptions. Accurate property value estimation is crucial for households, investors, financial institutions, regulators, and public policy. The goal is to avoid arbitrary measurements and provide reliable estimates for real estate value using appropriate tools.

2. Valuation Approaches

The market approach is widely regarded as the most appropriate method for valuing residential real estate in active markets. This approach relies on market efficiency and the similarity between properties. Hedonic models are often used to complete the sales comparison grid and establish a price index for properties with specific attributes. These models estimate the marginal effect of property attributes on sale prices based on observed data, providing more reliable valuations than subjective methods.

3. Regression Models vs. Artificial Neural Networks (ANN)

Various methods have been proposed for estimating property selling prices, including regression models and artificial intelligence techniques like ANN. ANN, despite being used since the 1980s, remains relevant due to advancements in big data and deep learning. ANN models are known for their superior predictive power, accuracy, and ability to handle large datasets. However, they are often criticized for their "black box" nature, which makes it difficult to interpret how input variables contribute to the output.

4. Comparative Studies

The literature indicates that ANN models generally outperform traditional regression models like Generalized Linear Models (GLM) in predictive accuracy. Studies have shown that ANN provides more precise outputs, better fits with statistical measures, and realistic pricing of individual properties. Despite this, regression models like GLM offer advantages in terms of transparency and interpretability, which are crucial for professional valuation environments.

5. Practical Applications and Implications

ANN models can serve as a valuable tool for reviewing valuation reports by providing a way to cross-verify evaluations done by professionals. They eliminate some subjectivity inherent in traditional valuation methods and ensure consistency in property value estimates. Given the performance metrics and practical results, ANN demonstrates better stability and predictive ability compared to GLM.

6. Future Directions

The ongoing development of optimization algorithms for ANN suggests that its performance in real estate valuation could improve further. There is potential for ANN to enhance automated valuation models and support more sophisticated property market analyses.


# **Own analyses**

Price performance analysis in real estate involves examining various attributes that influence property values. These attributes can be broadly categorized into measurable and observable factors, such as the living area size in square feet and the number of bedrooms, and subjective factors that require human interpretation, such as the view from the window. By integrating these diverse attributes into an AI-based analytical framework, we can systematically evaluate how each factor impacts property prices. This section outlines the parameters and methodologies used in our analysis, emphasizing the role of anti-discriminative optimizations and the significance of delta-values between initial and estimated prices. The data collection process for this analysis involved gathering information from various sources, including websites, APIs, and public datasets. While accessing data was relatively straightforward, ensuring its reliability posed significant challenges. Data from different companies or organizations often came with varying degrees of accuracy and consistency, and the use of different scaling systems for attributes like overall condition and quality of property view further complicated the analysis. Despite these challenges, we were able to compile a comprehensive dataset encompassing a wide range of attributes (e.g. Living area size in square feet, Number of bedrooms, Year property was built, Living area above ground level in square feet, Waterfront view, Number of bathrooms, Number of bathrooms)

Analytical Framework

To analyze the impact of these attributes on real estate prices, we utilized an online AI-tool capable of running anti-discriminative optimizations. This tool employs staircase functions to test the hypothesis regarding which attributes affect prices more significantly and which have a lesser impact. The main focus is to ensure that the model remains unbiased and equitable across different demographic and geographic groups.

Anti-Discriminative Optimizations

To ensure fairness in our price estimations, we incorporated anti-discriminative optimizations within our AI-tool. These optimizations help mitigate biases that might arise from subjective attributes and ensure that no particular demographic or geographic group is unfairly advantaged or disadvantaged. The staircase functions used in this process allow for a granular examination of each attribute's impact, providing a more nuanced understanding of property valuation.

**Rational and OAM**

These attributes provide a comprehensive overview of both property-specific and neighborhood-specific factors that can influence real estate prices. Property attributes include physical characteristics and qualitative assessments of the property itself, while neighborhood attributes account for the broader context and environment in which the property is located. Together, they form the basis for a thorough and accurate analysis of real estate price performance.

However, some attributes (e.g. Quality level of property view, and condition rating) involve human interpretation. Thus, it makes it difficult to gather data from different sources with consistent Furthermore it could affect the accuracy of the analysis so there should be a solution (e.g. software that can evaluate the quality of a property through image analyzation – see Figure#0):



**Figure#0 Various attributes of property and neighborhood of Real State, Source: Own presentation**

With the working model, attributes, objectives, and data at hand, the next step is to analyze this data using specialized tools such as AI and advanced data analytics. Initially, a rank analysis was performed using Excel features, followed by an AI analysis using COCO AI Analysis. Eighty-seven objects were randomly selected from a large open-source dataset for analysis. The target variable (Y) is the property price, while the other 19 attributes are the explanatory variables (X). The analysis aims to understand how these attributes influence property prices. Following attribute selection, Directions have been assigned to each feature. Attributes with higher values indicating a higher price were assigned a value of 0, while those with a negative correlation to price (higher value means lower price) received a value of 1. This process was further informed by the correlation analysis between each attribute and the target price. **See Figure #1** for a visual representation of the data and analysis process:



**Figure#1 Various Attributes (OAM) used for the Price Performance analysis, Source: Own presentation**

The rank feature (**See Figure #2**) arranged the test subjects according to their direction and, for each attribute, calculated which test subjects had the most significant impact on property prices, showing a stronger relationship. This helped us extract real and useful information from our raw data. However, we needed a more advanced tool to analyze the entire data set, considering all attribute rankings, and predict which test subjects and attributes are at higher risk.

For this task, we used a three-step process:

* Creating an auxiliary table using a Y (0) constant module.
* Calculating the scores and differences in the auxiliary table.
* Using online COCO analysis.

COCO Analysis is an AI-based analytical tool that helps predict pointers and key data directions in large data sets. This AI program can serve as a ranking module, identifying which attributes are more important and predicting outcomes. With data from COCO AI analysis, the data analytics team and company managers can leverage this information as a powerful strategy for data analytics and summarization.

The tool can be used by individuals and real estate agencies to determine the importance of each attribute, providing comprehensive measures and evaluations.



**Figure#2 Ranking Based on Excel Solver Module with conclusion after using, Source: Own presentation**

One of the most important tasks is to validate our evaluations to ensure they are correct and reliable. To achieve this, we reversed the initial ranked table and uploaded it to COCO STD again to assess the reliability of the data. (**See Figure#3**) Having reversed ranking table and its ratio we can determine if it is evaluable or not by multiplying two ratio results and comparing it with 0. With this our validation column created (**See Figure#2**):

****

**Figure#3 Reversed ranking table based on a Solver, Source: Own presentation (Legend: row-headers RS001-RS022)**

After obtaining comprehensive results from COCO AI and validating them, the filtered objects identified as potentially more expensive are shown (**See Figure#4**). In the table, objects RS024, RS061, and RS064 exhibit significantly higher delta values than the other objects, with differences exceeding $120,000.

 

**Figure#4 Delta values of the objects that could be more expensive, Source: Own presentation**

Conversely, object RS059, and RS067 identified as having less potential of reaching initial price, reaching positive delta value over $120000. **(See Figure#5)**

****

**Figure#5 Delta values of the objects that should be less expensive, Source: Own presentation**

It is possible that the initial and estimated prices could be almost the same after obtaining the estimated values. These occurrences **(See Figure#6)** show a ratio of approximately one percent and have remained remarkably constant.

 ****

**Figure#6 Objects that demonstrated less difference between initial and estimated price, Source: Own presentation**

The next step of the analysis was to determine the trend of the Delta values for different attributes to understand how various features influence the estimation process. Having more bedrooms and a waterfront location corresponds to higher prices. **(See Figure#7)** 

**Figure#7 Trends of Delta values (avg, min, max) vs Bedrooms, and waterfront, Source: Own presentation**



**Figure#8 Trends of Delta values (avg, min, max) vs number of floors, Source: Own presentation**

****

**Figure#9 Trends of Delta values (avg, min, max) vs number of bathrooms, Source: Own presentation**

Having a higher grade generally corresponds to a higher price, but in some scenarios, properties graded 5 and 7 were analyzed as more expensive. **(See Figure#10):**

****

**Figure#10 Trends of Delta values (avg, min, max) vs Overall grade of the property, Source: Own presentation**

****

**Figure#11 Trends of Delta values (avg, min, max) vs Quality of property’s view, Source: Own presentation**

The condition of the property has a lesser impact on price. This is evident from the average Delta values, where properties with condition levels 1 and 5 do not show significant differences in their corresponding prices. **(See Figure#12)**

****

**Figure#12 Trends of Delta values (avg, min, max) vs Overall condition of the property, Source: Own presentation**

# **Results**

The analysis aimed to evaluate the impact of various property and neighborhood attributes on real estate prices using an AI-based valuation model. The primary focus was on understanding how different features influence price estimations, with an emphasis on minimizing biases and ensuring accuracy.

Attribute Impact Analysis

The trend analysis (see Figure 7) indicates that properties with more bedrooms and those located on waterfronts generally command higher prices. The Delta values for these attributes show a clear positive correlation with property prices, reinforcing the significance of these features in the real estate market. A higher grade typically corresponds to a higher property price. However, an interesting anomaly was observed where properties graded 5 and 7 were found to be more expensive than anticipated (see Figure 10). This suggests that certain mid-tier properties might have unique features or market conditions driving their prices up. The condition of a property has a relatively minor impact on its price. The average Delta values for properties with condition levels 1 and 5 do not show significant differences (see Figure 12). This indicates that the overall condition, while important, may not be a primary driver of price variations in the analyzed dataset. Some properties showed minimal differences between initial and estimated prices (see Figure 6). This consistency highlights the effectiveness of the AI model in accurately predicting property values, with a ratio of approximately one percent between initial and estimated prices.

Validation and Reliability

The validation process involved reversing the initial ranked table and reassessing the data using the COCO STD tool (see Figure 3). This step ensured the reliability of the evaluations and confirmed the robustness of the AI model in handling diverse data attributes and market conditions.

The findings demonstrate that the AI-based valuation model effectively incorporates various property and neighborhood attributes to produce accurate and unbiased price estimations. Key attributes such as the number of bedrooms, waterfront location, property grade, and view quality significantly influence property prices. The lesser impact of property condition suggests that other factors may be more critical in determining market value. Overall, the model provides a comprehensive and reliable approach to real estate valuation, offering significant improvements over traditional manual methods.

# **Discussions**

Expanding real estate market

The real estate market is continually expanding as more individuals recognize the value of owning property and the long term benefit it offers. This growth, however, often brings challenges, such as biased decision-making and significant price fluctuations. Our objective is to develop software that serves as a data-driven, accurate alternative decision-making tool, benefiting anyone looking to make a lifetime investment in real estate.

Market challenges

Traditional methods of property valuation, characterized by manual appraisals and heuristic approaches, are susceptible to biases and inconsistencies. These methods can result in overly biased or significantly fluctuated prices, which may not accurately reflect the true value of properties. By leveraging AI and extensive datasets, our software aims to mitigate these issues, providing a more objective and consistent valuation process.

Constructive Feedback and future improvement

While our model has shown promising results, we acknowledge the importance of continuous improvement. Constructive feedback and criticism of our structure, OAM approach, analysis, and results are invaluable. Such feedback will not only help refine this project but also contribute to my personal and professional growth as an aspiring real estate expert. I encourage peers, mentors, and industry professionals to provide their insights and suggestions.

Contact for feedback and Suggestions

I invite you to reach out with your feedback and suggestions for this project. Your input is crucial for enhancing the effectiveness and accuracy of our AI-based valuation model. Please contact me at ssun41268@gmail.com with any comments or recommendations.

# **Conclusions and Summary**

In conclusion, the OAM matrix provides an excellent foundation for threat analysis, serving as a springboard for further analysis using AI. AI tools such as COCO analysis can handle complex datasets and predict outcomes to address real-world situations effectively.

The key takeaway is the value of Human-AI integration. By leveraging AI, humans can solve more complex problems in shorter timeframes, improving the overall efficiency, quality, and reliability of projects.

In this project, I, as a human creator, identified a real-world situation, developed an analysis framework for the problem, and curated the structures to address it. AI then assisted by analyzing the data from these structures, providing valuable insights and analysis that can be used to achieve comprehensive and reliable outcomes.

# **Future Directions**

Automation of Datasets

Automating dataset collection and preprocessing is crucial for enhancing the efficiency and scalability of the model. This will involve establishing automated connections with various real estate APIs to continuously import up-to-date data, implementing automated data cleaning and preprocessing routines to ensure data quality and consistency, and setting up real-time data monitoring systems to detect and incorporate changes in market dynamics instantly.

Use of Advanced AI Tools and Software

Incorporating more advanced AI tools and software can significantly enhance the accuracy and capabilities of the valuation model. This includes exploring and integrating advanced machine learning models, such as reinforcement learning, ensemble methods, and deep learning architectures, to improve prediction accuracy and robustness. Additionally, implementing explainable AI (XAI) techniques will make AI models more transparent and interpretable, increasing trust and usability among real estate professionals. Utilizing state-of-the-art optimization algorithms will help fine-tune model parameters and enhance performance, while leveraging big data technologies will enable the efficient handling and processing of large volumes of real estate data.

Fully Automated Process

Moving towards a fully automated valuation process involves seamless integration of data collection, preprocessing, analysis, and reporting into a cohesive, automated workflow. This includes developing an end-to-end automated system that manages every step of the valuation process, from data ingestion to final price estimation. Creating a unified workflow that integrates various AI tools, data processing modules, and reporting systems will ensure a smooth and efficient valuation process. Designing the system to be scalable and flexible will allow it to adapt to different market conditions and accommodate various property types and attributes.

Developing User-Friendly Applications

To make the valuation model accessible and practical for a broader audience, developing user-friendly applications is essential. This involves creating intuitive web and mobile applications that allow users to easily input property details and receive accurate price estimations. Designing interactive dashboards will provide users with visual insights and detailed analysis of property valuations. Offering customization options will enable users to adjust and refine parameters based on their specific needs and preferences. Focusing on user experience (UX) design principles will ensure the applications are easy to navigate, visually appealing, and provide a seamless user experience. Including educational resources and tutorials within the applications will help users understand how to use the tools effectively and interpret the results.

An automated working model, with appropriate attributes, would have a functional real-life application as it could be scaled up or down as per the needs of the company. Another addition to this project could be the assimilation of the project as a true AI-human integration project where human inputs are augmented by AI, and AI gets its logical direction and needs analysis from humans. This integration has begun in most industries of the world, and this project would focus on this direction too.

# **References**

1. Kauko, T., & d’Amato, M. (2008). Appraisal practices from an international perspective. –

[Kauko Tom and d’Amato Maurizio (Eds.): Mass Appraisal Methods. An international perspective for property valuers, 2008, Wiley-Blackwell, pp. 319, ISBN: 978-1-4051-8097-9 | Journal of Housing and the Built Environment (springer.com)](https://link.springer.com/article/10.1007/s10901-009-9156-y)

1. DiPasquale, D., & Wheaton, W. C. (1994). Hedonic price modeling in real estate. -

[Urban economics and real estate markets: DiPasquale, Denise: Free Download, Borrow, and Streaming: Internet Archive](https://archive.org/details/urbaneconomicsre0000dipa)

1. Worzala, E. et al. (1995). The market approach in residential real estate valuation. -

[An Exploration of Neural Networks and Its Application to Real Estate Valuation on JSTOR](https://www.jstor.org/stable/24881837)

1. Curry, B. et al. (2002). Advantages of hedonic price functions.

[Hedonic Price Model | SpringerLink](https://link.springer.com/referenceworkentry/10.1007/978-3-031-17299-1_1279)

1. Peterson, J. R., & Flanagan, A. B. (2009). Predictive precision of ANN in real estate.

[Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal by Steven P. Peterson, Albert B. Flanagan: SSRN](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1086702)

1. What Is an Automated Valuation Model (AVM)? How They Work –

[What Is an Automated Valuation Model (AVM)? How They Work (investopedia.com)](https://www.investopedia.com/terms/a/automated-valuation-model.asp)

1. [Top 7 Best Real Estate APIs (2022) [24+ Reviewed] (rapidapi.com)](https://rapidapi.com/blog/best-real-estate-apis/)
2. Artificial intelligence is taking over real estate – here’s what that means for homebuyers –

[What artificial intelligence means for homebuyers, real estate market (cnbc.com)](https://www.cnbc.com/2021/09/17/what-artificial-intelligence-means-for-homebuyers-real-estate-market.html)

1. COCO-online-tool: <https://miau.my-x.hu/myx-free/index_en.php3>