Literature Review of Explainable Machine Learning in Real Estate

Arnis Staško[1] and Jānis Grundspeņķis[1]

Riga Technical University, 6A Kipsalas Street, Riga, LV-1048, Latvia

Abstract

A literature review is conducted on explainable machine learning methods used in real estate. It identifies 17 relevant articles that reveal various subfields of real estate and the explainable machine learning methods used. Among them, XGBoost and SHAP is the most commonly used combination for explainable machine learning in the studied area. The study also identifies research gaps that could be addressed through further studies on time factors, model explainability, training set balance, and causal dependencies.

Keywords

Real estate, explainable machine learning, research methods, literature review

1. Introduction

The demand for artificial intelligence applications has grown significantly in the last decade. Companies are looking for ways to integrate artificial intelligence solutions into their processes to improve their product or service and competitiveness in the market, as well as to reduce the required amount of labour or costs. Real estate companies are no exception. There is a shortage of labor and customers expect lower operating costs under competitive conditions. It is essential to make the right decisions about real estate and its management where the number of influencing factors is large and difficult for a person to grasp. Therefore, artificial intelligence solutions could help.

Artificial intelligence studies methods for developing intelligent machines or software that imitate human behaviour. Although people usually talk about the need to implement an artificial intelligence solution, in practice it often results in the development of machine learning solutions. Machine learning is a subfield of artificial intelligence that creates software models from training examples to perform prediction, recognition, or clustering. Diverse machine learning algorithms allow us to train systems so that they gain autonomy, but the disadvantage of the most common ones based on neural networks is the inability to explain the obtained result (black box). Therefore, there is an increased interest in explainable machine learning methods, which would not only provide predictions or

Baltic DB&IS Conference Forum and Doctoral Consortium 2024
arnis.stasko@rtu.lv (A.Staško); janis.grundspenkis@rtu.lv (J.Grundspeņķis)
0000-0003-2876-5497 (A.Staško); 0000-0003-2526-4662 (J.Grundspeņķis)

recommend decisions but would also argue for the recommended solution (white box). In the field of real estate people are not ready to blindly trust artificial intelligence to make a decision about the most expensive thing they own. Explainability is therefore critical.

Therefore, the questions of this research are related to the need to investigate in which areas of real estate machine learning is used, what research methods and algorithms are used, why explainable machine learning is chosen and what further research might be useful. Accordingly, the research object is explainable machine learning methods.

The structure of the work is as follows. Chapter 2 explores the types of literature review used in similar studies. Further, Chapter 3 describes the literature review approach. Then, the literature review results are presented in Chapter 4. Finally, the conclusions and future work are summarized in Chapter 5.

2. Method Selection for Literature Review

To choose a suitable literature review method for the research, a search for publications in the ScienceDirect[2] database is carried out by searching ("machine learning" AND "literature review") in article titles and limiting results to 2023. Journal articles from the last year should be sufficient to reasonably identify the most current approaches. From 27 returned articles only 24 are used due to availability or title relevance.

Briefly browsing the content of the articles and paying special attention to the research method section, it is found that 20 out of 24 use systematic literature review. On closer examination, it is seen that the majority leans towards Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [20] or in the direction of Kitchenham and Brereton's various modifications of systematic review [12].

Considering that Kitchenham and Brereton's [12] specializes in software engineering literature reviews, while PRISMA guidelines [20] originate from the medical field, within the scope of this study the Kitchenham and Brereton's version [12] is adopted. The next chapter describes the approach of a literature review.

3. Literature Review Protocol

The literature review adapted from Kitchenham and Brereton's version [12] is performed as follows:

1. Define research questions for the literature review.
2. Perform an **initial search** in the ScienceDirect database by searching for review articles related to research questions to ensure that a similar literature review has not already been conducted.
3. Perform a **manual search** in the ScienceDirect database by searching for articles related to research questions. Select candidate papers based on abstract & title.
4. Iteratively perform **forward and backward snowballing** in the Scopus abstract and citation database[3]. Add any missed papers based on abstract & title analysis.

[2] https://www.sciencedirect.com/ (accessed December 21, 2023)
[3] https://www.scopus.com/ (accessed January 6, 2024)

5. Read the full version of selected papers and apply detailed inclusion/exclusion criteria during the **data extraction** and quality assessment process.

The authors believe that the use of the combination of ScienceDirect and Scopus provides sufficient coverage of reliable literature sources.

3.1. Research Questions

The cornerstone of a systematic literature review is the definition of research questions. So, to achieve the goals set for the research, the research questions are:

- RQ1. In what subfields of real estate explainable machine learning is applied?
- RQ2. What research methods are used to study explainable machine learning in the field of real estate?
- RQ3. What machine learning methods are used in the field of real estate?
- RQ4. What explainable machine learning methods are used in the field of real estate?
- RQ5. Why explainable machine learning methods are used in the field of real estate?
- RQ6. What are the research gaps in explainable machine learning in the field of real estate?

Further, the results of the availability of similar studies in the literature are analyzed.

3.2. Initial Search

To ensure that a similar reliable literature review is not available, ScienceDirect[4] is searched for keywords related to the research. Results for a search within article titles, an abstract and keywords are summarized in Table 1.

Table 1
Initial search results

Search phrase	Results
("real estate" AND "explainable machine learning" AND "overview")	0
("real estate" AND "explainable machine learning" AND "review")	0
("real estate" AND "explainable machine learning" AND "survey")	0
("real estate" AND "explainable artificial intelligence" AND "overview")	0
("real estate" AND "explainable artificial intelligence" AND "review")	0
("real estate" AND "explainable artificial intelligence" AND "survey")	0

4 https://www.sciencedirect.com/ (accessed January 6 , 2024)

The initial search results prove that a potentially similar literature review is not available. It is justified to carry out the intended literature review. Next, manual search results are summarized.

3.3. Manual Search

The manual search is performed in the ScienceDirect[5] database by searching for research articles by phrase ("real estate" AND ("explainable machine learning" OR "explainable artificial intelligence" OR “XAI”)) in article titles, an abstract and keywords. A total of five articles are found [1], [8], [10], [13], [19]. After reading the title and abstract, all are accepted as relevant for further research. If there are other publications, authors trust that they will be discovered in the process of snowballing in the Scopus database.

3.4. Forward & Backward Snowballing

In the forward snowballing all articles citing the examined article and in the backward snowballing all articles referenced from the examined article according to the Scopus[5] database are reviewed and the relevant articles are selected.

In the first iteration, the articles found during manual search are examined. In every next iteration, the articles found in the previous iteration are examined. As relevant are accepted articles between 2019 and 2023 with full-text availability and whose title or abstract reflects a connection with the field of real estate and use explainable machine learning methods in their research. A total of three iterations are performed. During the 3rd iteration, no new articles are found and the snowballing is not continued. The summary of all iterations and results is given in Table 2. With snowballing 12 new articles are added to the research.

Table 2
Summary of newly discovered and relevant articles discovered during forward and backward snowballing

Iter.	Source articles	Forward snowballing	Backward snowballing	Total
1	5	4: [2], [6], [15], [17]	6: [3], [9], [14], [16], [21], [22]	10
2	10	1: [5]	1: [23]	2
3	2	0	0	0

Once a list of relevant articles for further research is obtained, during the data extraction step the quality of articles is evaluated in detail and the answers to the research questions are clarified.

3.5. Data Extraction

According to the research questions data extraction and quality assessment are performed by reading the full text of each article. While the answers to RQ1, RQ3 and RQ4 are readily

[5] https://www.scopus.com/ (accessed January 7, 2024)

apparent, RQ2, RQ5 and RQ6 require additional effort. Almost none of the articles mention the exact research method used. In some of them a case study [3], [9], [10], [14] or a literature review [2], [8], [22] is mentioned, however, when researched in detail, it can be seen that the prime research method is a laboratory experiment. Similarly, the justification of the need for machine learning is to be explained. Several articles take this for granted and the detailed analysis of the benefits of explainability is performed to determine the real need. The most difficult is to determine research gaps. Therefore, the future research questions mentioned in the article are identified. Then, the actual research gaps are discussed. The data extraction results are presented in **Appendix A**.

4. Results

The literature review discovered 17 publications from scientific journals with Scopus cite scores between 3.3 and 14.8 (2023 data updated on 05.01.2024.). While the journal Habitat International[6] is ranked first in terms of the number of articles, the journal Reliability Engineering and System Safety[7] have the highest citation score 14.8. The full journal list is presented in Table 3. These results show that all articles are published in acknowledged editions.

Table 3
Journals presenting discovered articles

Journal	Cite Score 2023	Articles
Reliability Engineering and System Safety	14.8	1
Land Use Policy	13.3	1
Expert Systems with Applications	13.2	2
Finance Research Letters	10.8	2
Habitat International	10.2	3
Applied Geography	7.8	1
Big Data Research	7.8	1
Sensors	6.9	1
International Journal of Geo-Information (ISPRS)	6.7	1
Real Estate Economics	4.0	1
Journal of Real Estate Finance and Economics	3.7	1
Risks	3.6	1
Buildings	3.3	1

The literature study identified 64 authors publishing on the application of explainable machine learning in real estate. In terms of citations, the top most significant are the works of Kang & Zhang et.al. [9] with 86 citations, Chen & Yao et.al. [3] with 39 citations and Rico-

[6] https://www.sciencedirect.com/journal/habitat-international
[7] https://www.sciencedirect.com/journal/reliability-engineering-and-system-safety

Juan & Taltavull de La Paz [21] with 38 citations. Visualization is used to demonstrate the scope of the authors' contribution (**Figure 1**).

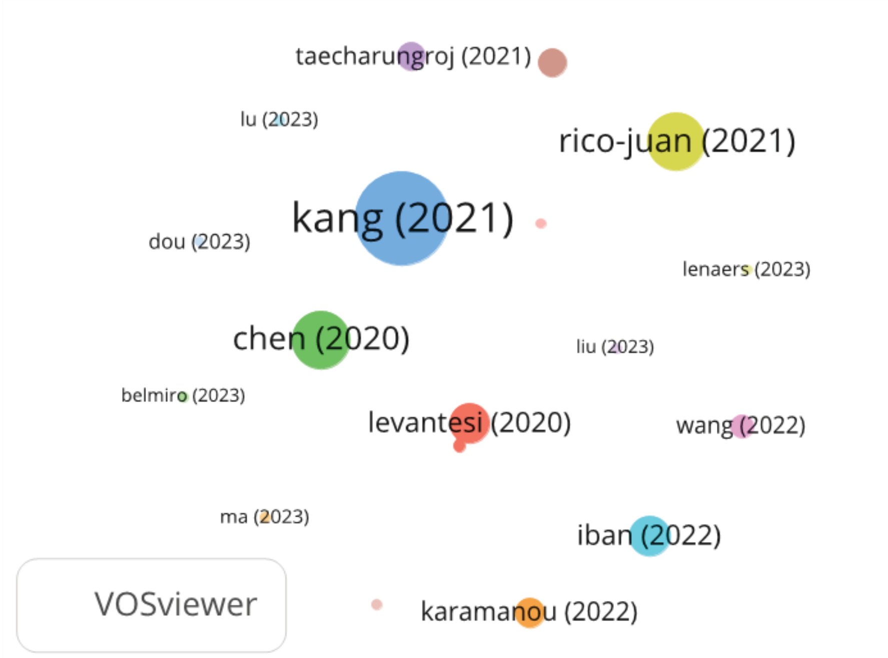

Figure 1: Author work by citations.

Significant to discover a set of keywords that illustrate the topic of the reviewed articles (**Figure 2**). They represent the research area.

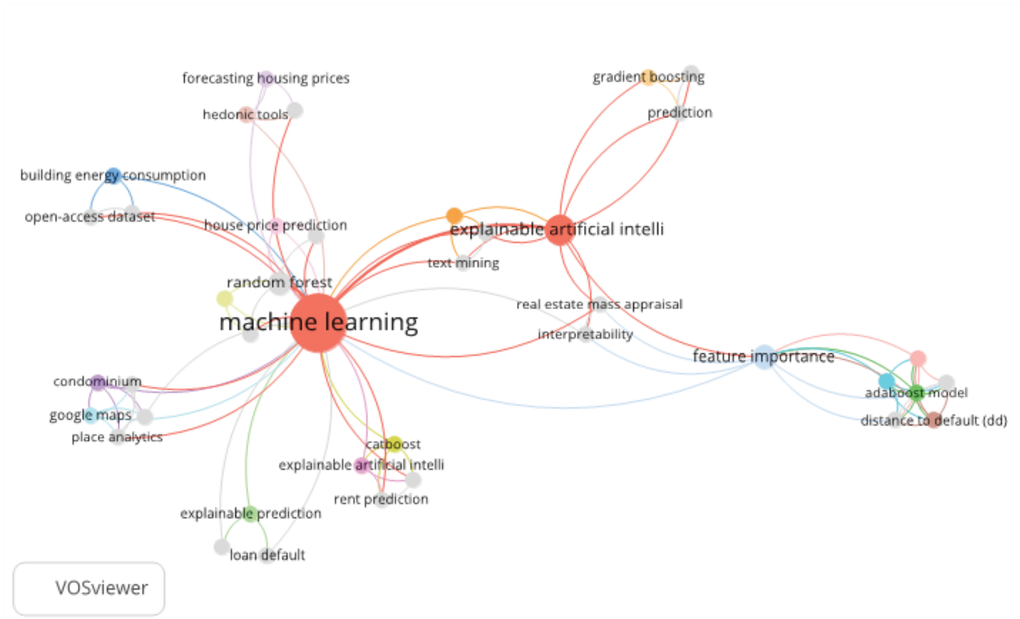

Figure 2: Keywords presenting discovered articles.

The subsections below summarize the answers to the research questions.

4.1. RQ1: Real estate subfields

The first research question RQ1 is "In what subfields of real estate explainable machine learning is applied?" The literature study reveals 8 different research subfields in real estate, where the most frequently addressed issue is real estate **price prediction** [1], [3], [8], [9], [10], [14], [21], [22], then follows real estate **price estimation** [16], [5] and real estate **rent price** prediction [6], [13]. One study from each subfield represents on understanding of the **land use intensity** [2], real estate **fire loss prediction** [23], building **thermal comfort requirement prediction** [15], stadium **fire risk assessment** [17] and **credit default prediction** of real estate companies [19]. This information gives an idea in which areas it would be possible to repeat similar studies in a reader's region, and also allows to navigate which directions have not yet been covered, in case new research is implemented.

4.2. RQ2: Research methods

The second research question RQ2 is "What research methods are used to study explainable machine learning in the field of real estate?" Evaluating all articles, it can be concluded that they all represent a laboratory experiment as a research method. This is quite understandable since building a machine learning model consists of training a model and evaluating its results using a testing set. Such an approach by default involves a laboratory experiment.

In addition, it should be noted that in four articles it is mentioned that a case study is conducted [3], [9], [10], [14]. On the other hand, from the content of three articles, it is observable that a literature review is carried out [2], [8], [22].

4.3. RQ3: Machine learning methods

The third research question RQ3 is "What machine learning methods are used in the field of real estate?" When searching for answers to this question, two aspects were evaluated - firstly, which machine learning methods are used and secondly, which of them shows the highest results or is the only one tested. The list of the machine learning methods studied in real estate is provided in Table 4.

The XGBoost method shows the best results or is chosen as appropriate in 7 out of 10 cases [3], [5], [6], [8], [10], [16], [22]. It is followed by Random forest in 4 out of 10 cases [2], [14], [17], [21] and LightGBM in 2 out of 4 cases [1], [15]. One in each study IBTEM [23], CatBoost [13], AdaBoost [19] and Gradient boosting machine [9]. The top three methods – XGBoost [4], Random Forest [7] & LightGBM [11] are based on decision tree algorithms. The results are useful as they allow to make research-based choices about the machine learning method for similar research.

Table 4
Machine learning methods studied in real estate

No	Method	Count	No	Method	Count
1	**XGBoost (#1)**	10	12	EBM	1
2	**Random Forest (#2)**	10	13	Elastic net	1
3	**LightGBM (#3)**	4	14	GBDT	1
4	AdaBoost	3	15	GBR	1
5	KNN	3	16	IBTEM	1
6	Linear regression	3	17	Lasso regression	1
7	CatBoost	2	18	Logistic regression	1
8	Decision tree	2	19	Multiple linear regression	1
9	Gradient Boosting	2	20	Naïve Bayes	1
10	Ridge regression	2	21	Neural network (Multilayer perceptron)	1
11	SVR	2	22	SVM	1

4.4. RQ4: Explainable machine learning methods

The fourth research question RQ4 is "What explainable machine learning methods are used in the field of real estate?" In the field of explainable machine learning, six different methods are used in the literature – SHAP [1], [3], [6], [8], [10], [13], [15], [17], [19], [21], [23]; FI [2], [9], [14], [16], [22]; PDPs [13], [14], [16], [22]; PFI [5], [8], [13]; ALE plots [2], [13], [16]; ICE [19]. The SHAP [18] method and its various modifications are the most widely used. The SHAP global and local explanations provide an opportunity to explain black box machine learning techniques. It allows to build a complex / black-box machine learning model that provides the highest possible results, while maintaining the possibility of understanding its operation, as well as gaining knowledge about the field under study.

4.5. RQ5: The reason for explainable machine learning

The fifth research question RQ5 is "Why explainable machine learning methods are used in the field of real estate?" Analyzing the publications, the reasons why their authors chose to use explainable machine learning methods can be interpreted in different ways, however, in fact, all researches found in the field of real estate are united by one goal - to understand the decision or forecast suggested by the model or to find correlations between the known information and the predicted outcome. Explainability simultaneously provides both knowledge of the researched field and increases users' confidence in the obtained solution. A detailed analysis can be found in **Appendix A**.

4.6. RQ6: Research gaps

The sixth research question RQ6 is "What are the research gaps in explainable machine learning in the field of real estate?" This is the most difficult question to analyze when studying the literature. The authors of each article indicate possible further work or improvements as a continuation of their research. However, that does not always indicate research gaps in general.

11 studies out of 17 note the need to repeat the study with better quality, additional or different types of data [1], [3], [5], [8], [9], [13], [14], [15], [17], [22], [23]. 8 studies note the need to improve the performance of algorithms by tuning them or testing others [1], [5], [9], [13], [14], [16], [21], [22]. 6 studies propose to try the solution in a different geographical location [2], [3], [6], [8], [14], [22]. 4 studies encourage to try a solution in real life or explore specific aspects of real life [2], [6], [19], [23]. 3 studies suggest improving the speed of the algorithm [5], [16], [17], or including the time factor [3], [9], [23] in the analysis of the problem sphere. Only 2 studies suggest improving model explainability [9], [21]. In conclusion, one study at a time encourages comparing the results of different fields [3], solving the imbalance of the data set [17] or looking for the true causal dependencies [5].

5. Conclusions

From the conducted literature review it is evident that explainable machine learning methods in the field of real estate are used to determine property value, rent and price, as well as land use intensity, fire damage, thermal comfort, fire risk and bankruptcy prediction.

In the field of machine learning, the most suitable research method is a laboratory experiment, and it is useful to apply a literature review and/or case study, if necessary. The study also indicates that the decision tree based XGBoost, Random Forest & LightGBM machine learning methods and SHAP explainable machine learning method are the most suitable or most used in real estate, providing the results of the highest value. The use of explainable machine learning is mainly necessary to understand the decision or forecast. Moreover, it provides an understating about the researched field and increases trust in the obtained machine learning model.

On the other hand, the study of research gaps gives only general ideas for further research. It's offered to make common improvements to existing solutions, to use additional data, to replicate the experiment in other areas or to try the solution in real-life situations. Scientific innovations could be sought in studies of time factors, model explainability, training set balance, and causal dependencies. However, before starting further research in these directions, additional research is needed to clarify what is done in specific technical areas that are not limited to real estate.

The results of this literature review can be used for further decisions on the implementation of similar research in the reader's region or for the initiation of new / unexplored research directions in the field of real estate.

Acknowledgements

The research leading to these results is part of the research project "Multi-contextual data analytics solutions for building management" jointly implemented by Riga Technical University, SIA "Lursoft IT" and SIA "Hagberg".

References

[1] Baur K., Rosenfelder M. & Lutz B., “Automated real estate valuation with machine learning models using property descriptions,” Expert Systems with Applications, vol. 213, pp.1-13, Mar. 2023.

[2] Belmiro C., Silveira Neto R.D.M., Barros A. & Ospina R., “Understanding the land use intensity of residential buildings in Brazil: An ensemble machine learning approach,” Habitat International, vol. 139, pp.1-12., Sep. 2023.

[3] Chen L., Yao X., Liu Y., Zhu Y., Chen W., Zhao X. & Chi T., “Measuring impacts of urban environmental elements on housing prices based on multisource data – a case study of Shanghai, China,” International Journal of Geo-Information (ISPRS), vol. 9 (no. 2), pp.1-23, Feb. 2020.

[4] Chen T. & Guestrin C., “XGBoost: A Scalable Tree Boosting System,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’16), 2016, pp. 785–794.

[5] Deppner, J., von Ahlefeldt-Dehn, B., Beracha, E. & Schaefers, W., “Boosting the Accuracy of Commercial Real Estate Appraisals: An Interpretable Machine Learning Approach,” Journal of Real Estate Finance and Economics, pp.1-38, Mar. 2023.

[6] Dou M., Gu Y. & Fan H., “Incorporating neighborhoods with explainable artificial intelligence for modeling fine-scale housing prices,” Applied Geography, vol. 158, pp.1-11, Sep. 2023.

[7] Ho T.K., “Random Decision Forests,” in Proceedings of the 3rd International Conference on Document Analysis and Recognition (ICDAR’95), 1995, pp. 278–282.

[8] Iban, M.C, “An explainable model for the mass appraisal of residences: The application of tree-based Machine Learning algorithms and interpretation of value determinants,” Habitat International, vol. 128, pp.1-11, Oct. 2022.

[9] Kang Y., F. Zhang F., Peng W., Gao S., Rao J., Duarte F. & Ratti C., “Understanding house price appreciation using multi-source big geo-data and machine learning,” Land Use Policy, vol. 111, pp.1-11, Dec. 2021.

[10] Karamanou A., Kalampokis E. & Tarabanis K., “Linked Open Government Data to Predict and Explain House Prices: The Case of Scottish Statistics Portal,” Big Data Research, vol. 30, pp.1-15, Nov. 2022.

[11] Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q. & Liu T.Y., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS'17), 2017, pp. 3148-3156.

[12] Kitchenham B. & Brereton P., “A systematic review of systematic review process research in software engineering,” Information and Software Technology, vol. 55 (no. 12), pp. 2049-2075, Dec. 2013.

[13] Lenaers I. & De Moor L., “Exploring XAI techniques for enhancing model transparency and interpretability in real estate rent prediction: A comparative study,” Finance Research Letters, vol. 58, pp.1-9, Dec. 2023.

[14] Levantesi S. & Piscopo G., “The Importance of Economic Variables on London Real Estate Market: A Random Forest Approach,” Risks, vol. 8 (no. 4), pp.1-17, Dec. 2020.

[15] Liu H. & Ma E., "An Explainable Evaluation Model for Building Thermal Comfort in China," Buildings, vol. 13 (no. 12), pp.1-20, Dec. 2023.
[16] Lorenz F., Willwersch J., Cajias M. & Fuerst F., "Interpretable machine learning for real estate market analysis," Real Estate Economics, vol. 51 (no. 5), pp. 1178-1208, Sep. 2023.
[17] Lu Y., Fan X., Zhang Y., Wang Y. & Jiang X., "Machine Learning Models Using SHapley Additive exPlanation for Fire Risk Assessment Mode and Effects Analysis of Stadiums," Sensors, vol. 23 (no. 4), pp.1-19, Feb. 2023.
[18] Lundberg S.M & Lee S.I., "A Unified Approach to Interpreting Model Predictions," in Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS'17), 2017, pp. 4768–4777.
[19] Ma Y., Zhang P., Duan S. & Zhang T., "Credit default prediction of Chinese real estate listed companies based on explainable machine learning," Finance Research Letters, vol. 58, Dec. 2023.
[20] Moher D., Liberati A., Tetzlaff J., Altman D.G., Antes G., Atkins D., Barbour V., Barrowman N., Berlin J.A., Clark J., Clarke M., Cook D., D'Amico R., Deeks J.J., Devereaux P.J., Dickersin K., Egger M., Ernst E., Gøtzsche P.C., Grimshaw J., Guyatt G., Higgins J., Ioannidis J.P.A., Kleijnen J., Lang T., Magrini N., McNamee D., Moja L., Mulrow C., Napoli M., Oxman A., Pham B., Rennie D., Sampson M., Schulz K.F., Shekelle P.G., Tovey D., Tugwell P., "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," PLoS Medicine, vol. 6 (no. 7), pp.1-6, Jul. 2009.
[21] Rico-Juan J.R. & Taltavull de La Paz P., "Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain," Expert Systems with Applications, vol. 171, pp.1-14, Jun. 2021.
[22] Taecharungroj V., "Google Maps amenities and condominium prices: Investigating the effects and relationships using machine learning," Habitat International, vol. 118, pp.1-12, Dec. 2021.
[23] Wang N., Xu Y. & Wang S., "Interpretable boosting tree ensemble method for multisource building fire loss prediction," Reliability Engineering and System Safety, vol. 225, pp.1-17, Sep. 2022.

A. Data extraction and assessment results

ID/ REF	RQ1: Subfields	RQ2: Research Method	RQ3: Machine Learning Method	RQ4: Explainable Method	RQ5: Why explain	RQ6: Gaps
A1 [8]	Real estate price prediction	Literature review Lab experiment	XGBoost (Random Forest, XGBoost, LightGBM, Gradient Boosting)	PFI, SHAP	To identify what the model takes into account when estimating real estate prices	1. Test the model in other cities; 2. Repeat the experiment with verified and reliable value data; 3. Repeat the experiment with the addition of socio-economic and demographic data.
A2 [10]	Real estate price prediction	Case study, Lab experiment	XGBoost	SHAP	1. To justify the reliability of a predictive model; 2. To understand which factors affect and determine the prices of houses.	n/a
A3 [1]	Real estate price prediction	Lab experiment	LightGBM (Linear regression, elastic net, SVR, random forest, LightGBM)	SHAP	To understand how the model arrives at decisions	1. Validate the influence of different descriptions on real estate price in a controlled laboratory experiment; 2. Prove that the difference between non-contextualized methods and contextualized embeddings increases even more through fine-tuning a pre-trained BERT model. 3. Repeat the experiment on real estate descriptions in other languages than English and German. 4. Extend the approach to the textual descriptions of short-term rent offers like hotel rooms or AirBnB offers.
A4 [13]	Real estate rent price prediction	Lab experiment	CatBoost (Ridge regression, XGBoost, CatBoost)	ALE plots, PDPs, PFI, SHAP	To gain a comprehensive understanding of the factors driving rent	Validate methodology on other real estate or financial-economic datasets and models to deepen our understanding of the substitutability, complementarity, benefits, and limitations of XAI techniques in finance
A5 [19]	Credit default	Lab experiment	AdaBoost (AdaBoost, EBM,	ICE, SHAP	To clearly understand	Implement results for practical applications.

ID/ REF	RQ1: Subfields	RQ2: Research Method	RQ3: Machine Learning Method	RQ4: Explainable Method	RQ5: Why explain	RQ6: Gaps
	prediction of real estate companies		Logistic regression, Random forest, SVM)		the ranking of feature importance and the impact on the prediction results	
A6 [2]	Understand the land use intensity	Literature review Lab experiment	Random forest (Random forest, XGBoost).	ALE plots, FI	To understand the factors responsible for the higher urban land use intensity in cities	1. Consider other urban realities; 2. Apply the model on commercial lots and its variation according to economic activities; 3. Investigate urban physical structure of urban centers.
A7 [6]	Real estate rent price prediction	Lab experiment	XGBoost	SHAP	To understand the relationships between housing units and their neighbourhoods	1. Test the model in other cities; 2. Analyze neighbourhood characteristic interactive or synergetic impacts on housing prices.
A8 [15]	Building thermal comfort requirement prediction	Lab experiment	LightGBM (Bayesian-optimized LightGBM, KNN, Random forest, XGBoost, GBDT, SVR)	SHAP	To understand the thermal requirements of building occupants	Incorporate additional variables in the model
A9 [17]	Stadium fire risk assessment	Lab experiment	Random forest (Naïve Bayes, KNN, Decision tree, AdaBoost, LightGBM, Random forest)	SHAP	To find the complex nonlinear relationship between risk features and stadium fire risk.	1. Repeat the experiment with additional data; 2. Explore ways to solve the label imbalance; 3. Increase operational efficiency and reduce time costs;
A10 [3]	Real estate price prediction	Case study Lab experiment	XGBoost (Linear Regression, XGBoost, Random forest, GBR)	SHAP	To explain the impacts of urban environmental elements on housing prices	1. Test the model in other cities; 2. Quantify the differences among cities; 3. Integrate multi-year data to analyze the temporal dynamics of the impacts of the urban environmental elements on housing prices; 4. Repeat the experiment with additional and improved data.

ID/ REF	RQ1: Subfields	RQ2: Research Method	RQ3: Machine Learning Method	RQ4: Explainable Method	RQ5: Why explain	RQ6: Gaps
A11 [22]	Real estate price prediction	Literature review Lab experiment	XGBoost (Random forest, XGBoost)	FI, PDPs	To investigate the relationship between neighborhood amenities and the prices of condominiums	1. Repeat the experiment with additional data; 2. Additionally test and tune the model; 3. Identifying the similarities and differences in the importance of amenities in various geographical areas.
A12 [14]	Real estate price prediction	Case study Lab experiment	Random forest	FI, PDPs	To better explain which variables have more importance in describing the evolution of the house price following an urban approach	1. Repeat the experiment with additional and improved data; 2. Repeat the experiment with different machine learning algorithms; 3. Repeat the experiment on other real estate datasets.
A13 [9]	Real estate price prediction	Case study Lab experiment	Gradient boosting machine (GBM) with decision trees (Gradient boosting machine (GBM), Multiple linear regression (MLR))	FI	To examine the effect of different variables on house price appreciation	1. Test the model in other cities; 2. Repeat the experiment with DCNN; 3. Include dynamics of urban land use changes into the framework with richer datasets; 4. Add deeper exploration and more explanations. 5. Involve more time-series data and approaches from the economy to build causality relationships and improve the interpretability of the model.
A14 [21]	Real estate price prediction	Lab experiment	Random forest (KNN, Decision tree, Random forest, AdaBoost, CatBoost, Neural network (Multilayer perceptron), Linear regression, Ridge regression, Lasso regression)	SHAP	To observe non-linear relationships between housing prices and housing attributes	1. Repeat the experiment with deep artificial neural networks; 2. Resolve explainability challenges in deep artificial neural networks.

ID/ REF	RQ1: Subfields	RQ2: Research Method	RQ3: Machine Learning Method	RQ4: Explainable Method	RQ5: Why explain	RQ6: Gaps
A15 [16]	Real estate price estimation	Lab experiment	XGBoost	ALE plots, FI, PDPs	To justify decisions and generate new insights	1. Enhance model building speed; 2. Improve the reliability and validity of algorithmic decision-making.
A16 [23]	Real estate fire loss prediction	Lab experiment	IBTEM (Catboost, XGBoost, LightGBM)	SHAP	To understand the reasons for making certain decisions or predictions	1. Use a model to assist relevant departments in making timelier decisions regarding dispatching aid and mobilizing resources; 2. Repeat the experiment with updated data; 3. Include time series forecasting in the building fire loss prediction.
A17 [5]	Real estate price estimation	Lab experiment	XGBoost	PFI	To make informed decisions	1. Improve data availability for machine learning experiments; 2. Justify the rationale behind patterns or determine causality in the relation between input and output data; 3. Enhance model building speed; 4. Improve machine learning algorithms for the field of real estate.