

## 10th International Conference on Science & Social Research

Virtual Conference

6 - 7 Nov 2023

Organised by: Research Nexus UiTM (ReNeU), Universiti Teknologi MARA

### Big Data Analytics in Real Estate Valuation: A systematic literature review

Nur Edrina Muhammad Nazrin<sup>1</sup>, Abdul Hadi Nawawi<sup>1</sup>,  
Aida Wati Zainan Abidin<sup>2\*</sup>, Edie Ezwan Mohd Safian<sup>3</sup>

*\*Corresponding Author*

<sup>1</sup> Faculty of Built Environment, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

<sup>2</sup> Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

<sup>3</sup> Faculty of Technology Management and Business, Universiti Tun Hussein Onn, 86400 Parit Raja, Johor, Malaysia

edrinanazrin@gmail.com, abdul274@uitm.edu.my, aida018@uitm.edu.my, edie@uthm.uitm.edu.my  
Tel: +60139554795

---

#### Abstract

This systematic literature review (SLR) assesses the applicability of big data analytics (BDA) in real estate valuation by examining 77 studies published between 2018 and 2023. The findings reveal that this is an emerging research area with a growing body of literature. Key application contexts identified are property price prediction, property rental prediction, land value prediction, and property price indexing, with takeaways suggesting that BDA has the potential to enhance the accuracy of property valuation. Building on these insights, this study develops a comprehensive framework to synthesise the existing literature and serves as a foundation for future researchers.

Keywords: Big data analytics; real estate; valuation, systematic literature review

eISSN: 2398-4287 © 2025. The Authors. Published for AMER by e-International Publishing House, Ltd., UK. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers) DOI: <https://doi.org/10.21834/e-bpj.v10iSI40.7700>

#### 1.0 Introduction

The real estate market is a complex and ever-changing landscape influenced by multiple factors, resulting in substantial price fluctuations that are challenging to predict. Accurate and efficient prediction of real estate prices is crucial, as it influences decision-making across the real estate market (Shbool et al., 2025). Recently, there has been an increase in the quantity and quality of information on property prices and characteristics, enabling the application of more sophisticated statistical techniques. Given the significant size and economic importance of real estate as an asset class, exploring the potential benefits of big data analytics (BDA) in its valuation is of utmost importance to enhance prediction accuracy and automate large-scale property valuation (Saral, 2024). This approach enables property prices to be estimated without requiring a valuer or appraiser to be physically present (Alfaro-Navarro et al., 2020).

The aim of this study is to explore the application and role of BDA in real estate valuation. To achieve this aim, this study has the following research objectives: (1) To identify the current status of research on the application of BDA in real estate valuation, (2) To examine the contexts within which BDA applications in real estate valuation are being studied, and (3) To synthesise key insights derived from prior research on BDA in this field. Corresponding to these objectives, the study addresses the following research questions: (1)

eISSN: 2398-4287 © 2025. The Authors. Published for AMER by e-International Publishing House, Ltd., UK. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers) DOI: <https://doi.org/10.21834/e-bpj.v10iSI40.7700>

What is the current status of research on the application of BDA in real estate valuation? (2) In which contexts are these applications being studied? (3) What are the key insights derived from prior research on BDA? By addressing these aspects, this study offers two main contributions: a comprehensive synthesis of the existing literature and a guiding framework to support future research in this area.

## 2.0 Definition of Key Concepts

### 2.1 Big Data and Big Data Analytics

Big data comprises large, complex datasets that are difficult for traditional tools to handle efficiently (Nei et al., 2023). Much of this data is generated by individuals and technologies (Azizi & Rudnytskyi, 2022). Big data analytics (BDA) entails extracting useful insights from these extensive datasets using advanced computational and statistical methods (Chai et al., 2021). It leverages various tools and strategies to analyse enormous volumes of complex data to improve performance. As a sophisticated technique, BDA processes data from large-scale databases to find relevant information. It uses advanced analytical methods to examine both structured and unstructured data from sources such as social media, sensors, and transactions. This approach draws valuable insights from diverse datasets, ultimately supporting better decision-making (Lifang et al., 2023).

### 2.2 Real Estate Valuation

Real estate valuation is a complex and vital process in the property market. It determines the current worth of future benefits from property ownership and is essential for purposes such as financing, investment analysis, insurance, taxation, and auctions (Folger & Mansa, 2023). The unique features of each property complicate the process. Valuation considers factors such as land costs, market trends, property specifics, and other relevant elements (Bhosale & Turate, 2025).

## 3.0 Methodology

### 3.1 Protocol

The protocol for the systematic literature review (SLR) comprises three sequential processes namely planning, performing, and presenting the review (Tandon et al., 2020). The SLR utilizes pre-determined inclusion and exclusion criteria to ensure the selection of relevant literature. This structured approach ensures transparency, replicability, and methodological rigor, which strengthens the reliability of the review findings.

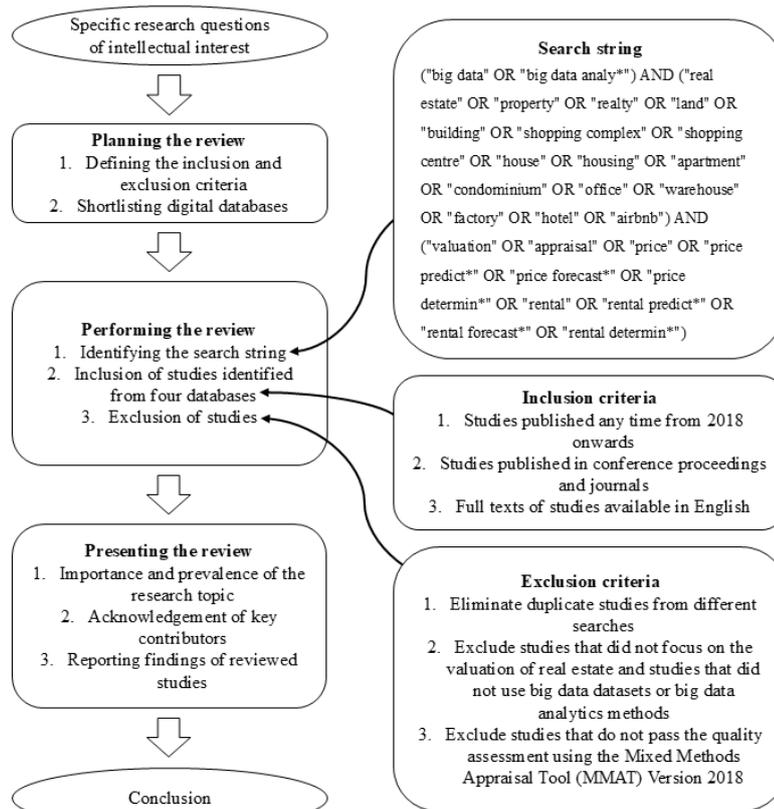


Fig. 1: Review Protocol

### 3.2 Planning the Review

The SLR examined four databases: Web of Science, Scopus, IEEE Xplore, and SpringerLink. Full texts of potentially relevant studies were carefully reviewed for eligibility. These databases were chosen to ensure broad coverage relevant to BDA in real estate valuation. Those meeting the inclusion and exclusion criteria were then assessed for quality and relevance to the research question. The requirements considered several factors: first, the article type—focusing on journal articles and conference proceedings while excluding non-research articles, book chapters, and series; second, only publications in English were included to avoid translation issues; and third, the publication date range was from 2018 to 2023. The review mainly focused on how big data is used in real estate valuation, with particular emphasis on studies from social science databases. Studies from any country were eligible. These criteria were applied to ensure methodological rigor, minimise bias, and capture recent and globally relevant studies. A detailed summary of these selection criteria is provided in Table 1.

Table 1. Criteria Selection

Criteria	Inclusion	Column B (t)
Article Type	Journal articles and conference proceedings	Non-research article
Language	English	Non-English
Year of Publication	2018 until 2023	<2018
Database	Web of Science, Scopus, IEEE Xplore, SpringerLink	None
Country	Any countries	None

### 3.3 Performing the Review

The search string used for the SLR was ("big data" OR "big data analy\*") AND ("real estate" OR "property" OR "realty" OR "land" OR "building" OR "shopping complex" OR "shopping centre" OR "house" OR "housing" OR "apartment" OR "condominium" OR "office" OR "warehouse" OR "factory" OR "hotel" OR "airbnb") AND ("valuation" OR "appraisal" OR "price" OR "price predict\*" OR "price forecast\*" OR "price determin\*" OR "rental" OR "rental predict\*" OR "rental forecast\*" OR "rental determin\*"). The search retrieved 1,268 studies. After reviewing titles and abstracts, 224 studies were selected for further screening. Duplicates from different databases were eliminated, leaving 159 unique studies. Full texts of these were then thoroughly examined, resulting in 91 studies that met the inclusion criteria. To ensure methodological quality and reliability, the Mixed Methods Appraisal Tool (MMAT), Version 2018, was used. The MMAT was chosen because it allows rigorous assessment across qualitative, quantitative, and mixed-methods studies, which is important given the interdisciplinary nature of BDA in real estate. Ultimately, 77 studies were included. A summary of the study selection process is shown in Table 2.

Table 2. Summary of Study Selection

Database	No. of Papers Retrieved	After Skimming Title and Abstract	After Excluding Duplicates	After Skimming Full Text	After Quality Assessment
Web of Science	514	102			
Scopus	461	74	64 excluded	77 excluded	5 excluded
IEEE Xplore	96	30			
SpringerLink	197	17			
Total	1,268	224	159	82	77

### 3.4 Presenting the Review

The SLR examines the evolution of research on BDA in real estate valuation, as illustrated in Fig. 2. Scholarly activity in this area emerged in 2018 with only six studies, indicating an initially limited research interest. This was followed by a substantial increase in 2020, when the number of publications increased to 13, signalling growing academic recognition of BDA's relevance to real estate valuation. Research output continued to expand thereafter, peaking at 28 studies in 2022, which reflects the consolidation of BDA as an important research domain within the real estate literature. Publications reported for 2023 represent studies published up to 30 June 2023 and are therefore lower than those recorded for preceding full year.

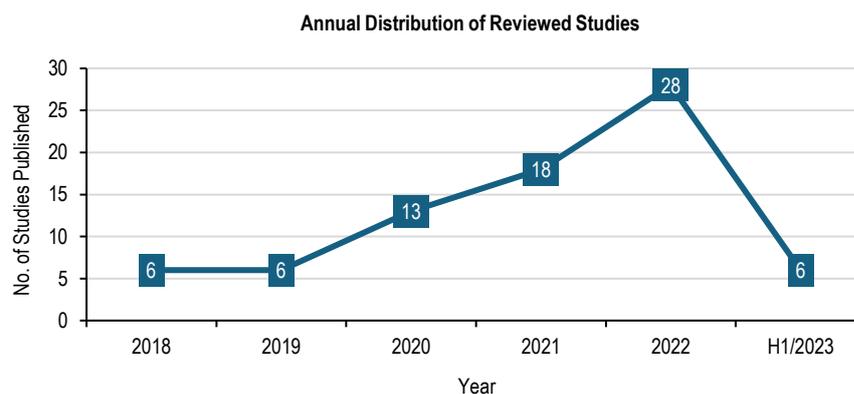


Fig. 2: Annual Distribution of Studies

## 4.0 Findings

### 4.1 Descriptive Analysis of the Literature

The analysis of the 86 selected studies reveals several key trends regarding the publication outlets, property types, and methodologies.

#### 4.1.1 Publication Trends and Outlets.

The sample comprised 61 journal articles and 25 conference proceedings. Among the four databases used in this study, Web of Science contributed the most studies, followed by Scopus, IEEE Xplore, and SpringerLink. The studies cover a wide range of property types, reflecting the diverse focus of real estate research within the SLR sample (see Fig. 3).

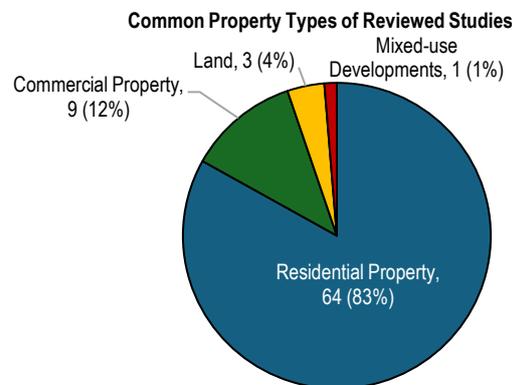


Fig. 3: Common Property Types of Reviewed Studies

#### 4.1.2 Common Property Types.

The property types examined in the SLR cover a broad spectrum, reflecting the diverse aspects of real estate research within the sample. Housing was the most analysed category, appearing 56 times, and includes various residential structures such as apartments, flats, detached, semi-detached, and terraced houses. Apartments, a subset of housing, appeared 13 times and play a unique role in urban environments, warranting specific attention. Commercial properties, including offices, retail spaces, and industrial sites, ranked third with five studies. Land was also explored in three studies, covering undeveloped areas and designated plots, emphasising its importance for urban planning and real estate development. Additionally, other property types, such as hotels, unique properties, Airbnb accommodations, and suites, were included to provide a more comprehensive view of real estate considerations.

#### 4.1.3 Common Methodologies.

The studies included in the SLR used various methodologies. Machine learning was the most common approach, with 38 studies employing techniques such as ensemble methods, including Gradient Boosting (GB), AdaBoost, and Stacked Random Forest Models, as well as specific algorithms like XGBoost, LightGBM, and CatBoost. Regression analysis was also frequently employed in 30 studies, utilising methods such as Ordinary Least Squares (OLS), Linear Regression, LASSO Regression, Elastic Net Regression, Huber Regression, RANSAC Regression, and Geographically Weighted Regression (GWR). Neural Networks appeared in five studies, including Back-Propagation (BP), Convolutional Neural Networks (CNNs), and Artificial Neural Networks (ANNs). A few studies incorporated Bayesian approaches and Spatial Analysis techniques, offering spatial insights and localised analyses.

Additionally, some studies explored innovative methods, including Geometric Brownian Motion, Monte Carlo Simulation, and Multimodal Learning. Various data analysis techniques, including exploratory factor analysis, correlation analysis, and cross-correlation analysis, were also utilised to deepen the understanding of data relationships. Notably, while regression analysis was commonly used in 30 studies, there is a clear trend showing that recent research increasingly favours ensemble machine learning techniques. This indicates a shift in the field towards more complex, non-linear models that better capture the intricate dynamics of the real estate market, as these have proven to offer superior predictive capabilities compared to traditional linear models.

### 4.2 Context of Big Data Analytics in Real Estate Studies

The contexts of the studies can be synthesised into four broad categories: (1) Property price prediction, (2) Property rental prediction, (3) Land value prediction, (4) Property price indexing (see Table 3).

Context	No. of studies	Details
Property price prediction	58	<ul style="list-style-type: none"> <li>• House price prediction</li> <li>• Commercial properties price prediction</li> <li>• Apartment price prediction</li> <li>• Flat and apartment price prediction</li> <li>• Apartment and house price prediction</li> </ul>

Property rental prediction	15	<ul style="list-style-type: none"> <li>• Residential properties price prediction</li> <li>• Residential and commercial properties price prediction</li> <li>• Apartment, condominium, and suite price prediction</li> <li>• Public and private buildings price prediction</li> <li>• Unique properties price prediction excluded</li> <li>• House rental prediction</li> <li>• Airbnb rental prediction</li> <li>• Apartment rental prediction</li> <li>• Residential properties rental prediction</li> <li>• Flat rental prediction</li> <li>• Hotel room rate prediction</li> </ul>
Land value prediction	3	<ul style="list-style-type: none"> <li>• Urban land value prediction</li> <li>• Benchmark land price prediction</li> <li>• Residential land value prediction</li> </ul>
Property price indexing	1	<ul style="list-style-type: none"> <li>• Apartment, industrial, office, and retail properties price indexing</li> </ul>

#### 4.2.1 Property price prediction.

Property price research has been a prominent topic in the real estate industry for many years. Property price prediction involves estimating the value of real estate properties. In the SLR, 58 distinct studies were identified that focus on predicting property prices. Within this scope, house price forecasting has been a primary focus.

Additionally, predicting commercial property prices has provided valuable insights. The study of flat and apartment price prediction is also significant, with some research exploring both property types. Residential property price forecasting and predicting both residential and commercial property prices, as discussed in various studies, are also prominent areas. Moreover, research extends to apartment, condominium, and suite price predictions, with contributions from studies on public and private building prices and the valuation of unique properties, thereby completing the diverse field of property price prediction research.

#### 4.2.2 Property rental prediction.

Research on property rental primarily examines how various factors affect the rental market, but less attention has been given to rental prediction. The SLR highlights 15 key studies, forming the second-largest research cluster in this area. House rental prediction has been a significant focus, while Airbnb rental prediction- a specialised segment- has also gained prominence. Similarly, interest in apartment rental forecasts has increased. Predicting rental rates for residential properties enhances the overall understanding of this market, including flat rentals, which are a specific segment. Hotel room rate prediction, however, remains a separate topic context.

#### 4.2.3 Land value prediction.

The SLR has identified three distinct studies that contribute to the context of land value prediction. These studies delve into predicting urban land values and contribute to the field by focusing on land price forecasts. They offer insights into benchmark land price prediction, which is essential for establishing reference points in the valuation of land assets and explore the prediction of residential land values.

#### 4.2.4 Property price indexing.

Having reliable real estate price indices is essential. Several methods have been proposed in the real estate literature to compile transaction price indices. Two of the most popular methods are hedonic pricing models and repeat sales. Hedonic pricing is the original method for constructing price indexes. However, BDA, particularly machine learning techniques, are now utilised to determine real estate price indices, especially in commercial properties such as apartments, industrial, office, and retail.

### 4.3 Value Delivered by Big Data Analytics in Real Estate Valuation

The capacity of BDA to significantly enhance real estate valuation becomes evident after a comprehensive review of the studies examined SLR.

#### 4.3.1 Data governance.

The use of diverse datasets in various real estate valuation studies in the SLR highlights several issues related to data governance. These datasets, included in the SLR, comprise both structured and unstructured data from multiple sources, including online real estate platforms, government agencies, social media, and others. Studies within the SLR demonstrate data governance by ensuring data privacy and security, thereby preventing the exposure of sensitive information, and protecting both organisations and individuals when necessary. They also employ data cleaning and preprocessing techniques to enhance data quality, handle missing values, and mitigate outliers. Additionally, some studies incorporate open data sources, supporting transparency and ethical data practices. The SLR emphasises the need for moral and legal compliance when handling large datasets and considerable data to maintain the accuracy and reliability of the information used in real estate valuation studies.

#### 4.3.2 Technological advancement.

The SLR highlights the transformative role of technology in revolutionising real estate valuation. The studies within the SLR significantly contribute to the field by illustrating how technology-driven innovations enable more accurate and efficient predictions of property prices, rentals, land value, and property price indexing. Furthermore, the SLR emphasises the shift from manual data collection to automated processes supported by innovative technologies, including web scraping, Python data retrieval, web crawling, geospatial data extraction,

web mining, and data mining techniques. This technological progress not only enhances the speed and accuracy of data collection but also equips real estate professionals with valuable insights for better decision-making, ultimately advancing the practice of real estate valuation in the digital age.

#### 4.3.3 Real estate prediction.

The SLR demonstrates how BDA revolutionises the real estate industry, enabling accurate predictions by leveraging a variety of data sources. These findings emphasise the crucial role of big data in improving prediction accuracy across various aspects of property valuation. The studies in the SLR have thoroughly tested a range of methods and algorithms, revealing key insights. Machine learning algorithms, particularly non-linear tree-based models such as decision trees, have demonstrated superior predictive capabilities compared to traditional linear models in forecasting housing values. Moreover, incorporating big data from real estate websites allows models to surpass conventional regression techniques, as shown by the superior performance of the random forest model in predicting apartment prices. Ensemble machine learning techniques, such as gradient boosting and random forest, have also demonstrated enhanced predictive capabilities. Notably, the significance of BDA lies in its success in predicting property prices, forecasting property rentals, estimating land values, and indexing property prices. These findings illustrate how big data, diverse data sources, and advanced analytics are driving a paradigm shift towards more precise and efficient real estate valuation. Insights from the SLR have contributed to a comprehensive framework encompassing key components of BDA in real estate valuation, as depicted in Fig. 4.

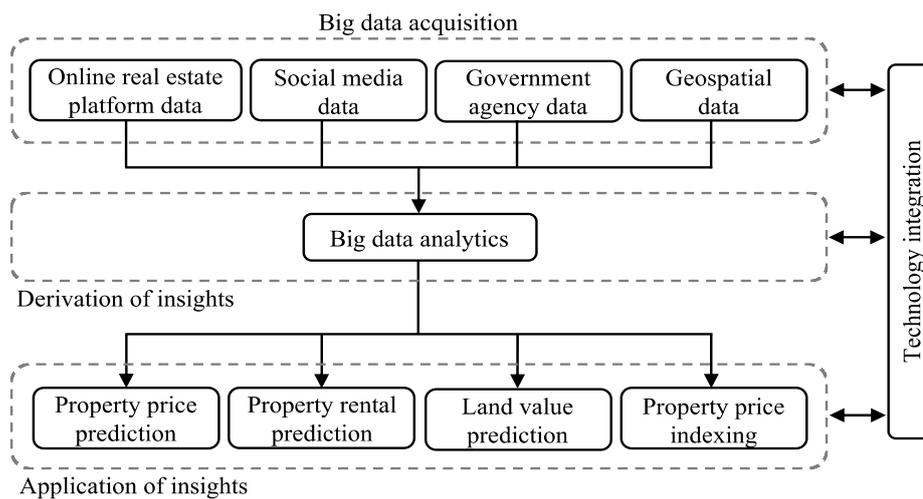


Fig. 4: BDA in Real Estate Valuation Framework

## 5.0 Discussion

### 5.1 Summary and Interpretation of Findings

The review confirms that the application of BDA in real estate valuation is a rapidly expanding field, with a notable rise in studies since 2018 and a peak in 2022. The research primarily focuses on property price prediction, with machine learning being the most commonly employed approach. This preference for machine learning, especially non-linear and ensemble models, reflects the complexity of the real estate market. These advanced models have demonstrated superior predictive capabilities compared to traditional linear methods, as they more effectively capture the complex factors that influence property values.

In contrast, property indexing remains notably under-explored, with reviews citing only one study that has utilised BDA techniques in this area. This highlights a significant research gap, as conventional methods, such as hedonic pricing, continue to dominate, despite the potential for machine learning to improve the precision of these critical economic indicators.

The application of advanced BDA tools, such as machine learning, regression analysis, neural networks, spatial analysis, Bayesian methods, multimodal learning, Monte Carlo simulation, and geometric Brownian motion, opens up exciting opportunities. These cutting-edge techniques could revolutionise the industry by enhancing the accuracy and efficiency of property valuation. Building on these findings and identified gaps, further research should examine how to incorporate these technologies into current real estate practices, thereby improving efficiency and precision and fostering innovation in valuation. Such initiatives could have a profound impact on the future landscape of real estate valuation in the digital age.

### 5.2 Implications of the Research

The findings of this SLR have important implications for both researchers and industry professionals. For academia, the clear focus on price prediction highlights an opportunity to diversify the research agenda. Future studies could offer significant value by exploring less-developed contexts, such as rental prediction and land valuation, or by applying BDA to new challenges, such as the valuation of sustainable or green properties. For industry professionals, the findings signal a fundamental shift in the practice of real estate valuation.

The proven success of BDA in improving prediction accuracy and enabling large-scale automated valuations points toward a future in which data-driven models supplement or even replace some traditional manual appraisal methods. This underscores the growing importance for valuers and real estate analysts to develop data analytics skills to remain competitive and to leverage these powerful new tools for informed decision-making.

### 5.3 Elaboration of the Proposed Framework

The framework developed from this review encapsulates the core components of applying BDA in real estate valuation. The three pillars, data governance, technological advancement, and real estate prediction-are deeply interconnected. Effective data governance serves as the foundation, ensuring that the large, diverse datasets used are high-quality, secure, and handled ethically. This foundation enables technological advancements, allowing automated processes such as web scraping and geospatial data extraction to collect vast amounts of data efficiently and accurately. Ultimately, this high-quality, large-scale data enables the development of real estate prediction models, enabling advanced machine learning algorithms to generate valuations with increased accuracy and precision. This framework offers a comprehensive view, illustrating that successful BDA applications encompass not only advanced algorithms but also the entire process, from data management to predictive output.

## 6.0 Conclusion and Recommendations

The SLR addressed three primary research questions regarding BDA in real estate valuation. It primarily focused on four key areas: (1) Property price prediction, (2) Property rental prediction, (3) Land value prediction, (4) Property price indexing. While it advanced understanding of the field, some limitations should be acknowledged. Excluding book chapters, magazine articles, and theses, and considering only English-language journal and conference publications, may have missed relevant insights. Additionally, relying on specific databases might have overlooked studies from other sources. Future research should address these limitations by expanding its scope for a more comprehensive overview. The use of advanced BDA technologies presents exciting opportunities for exploration. Based on the findings and gaps identified in the literature, future research should expand its scope to achieve a more comprehensive understanding. For example, studies could examine how innovative approaches, such as multimodal learning, can combine visual data (e.g., property photos) with traditional tabular data to enhance valuation accuracy. Another important direction is to investigate BDA applications in niche markets, such as valuing sustainable green buildings, an area that is notably missing from the reviewed literature.

## Acknowledgements

The authors acknowledge the Ministry of Higher Education (MOHE) for funding under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2019/SSI11/UITM/01/2).

## Paper Contribution to the Related Field of Study

This paper makes two significant contributions to the fields of real estate valuation and data analytics. First, it offers a thorough overview of the current literature on BDA in real estate valuation. By systematically reviewing 77 studies from 2018 to 2023, it maps out the research landscape, highlighting the field as emerging and categorising key application areas such as property price prediction, rental forecasting, and land value estimation. This synthesis offers researchers a well-structured and up-to-date overview of this rapidly evolving topic. Second, building on these insights, the study develops and proposes a guiding framework for future research. This framework combines the key components and processes found in the literature- such as data governance, technological advances, and prediction methods- into a cohesive model. It serves as a foundational guide for academics and practitioners, outlining a clear path for future studies in this domain.

## References

- Alfaro-Navarro, J. L., Cano, E. L., Alfaro-Cortés, E., García, N., Gámez, M., & Larraz, B. (2020). A Fully Automated Adjustment of Ensemble Methods in Machine Learning for Modeling Complex Real Estate Systems. *Complexity*, 1–12. <https://doi.org/10.1155/2020/5287263>
- Azizi, I., & Rudnytskyi, I. (2022). Improving Real Estate Rental Estimations with Visual Data. *Big Data and Cognitive Computing*, 6(3), 1–21. <https://doi.org/10.3390/bdcc6030096>
- Bhosale, P., & Turate, A. (2025). Review of Valuation Methods for Specialized Properties in the Real Estate Sector, *International Journal on Science and Technology (IJSAT)*, 16(2), 1–6. <https://doi.org/10.1108/PM-10-2024-0111>
- Chai, W., Labbe, M., & Stedman, C. (2021). Big Data Analytics. *TechTarget*. <https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics>
- Folger, J., & Mansa, J. (2023). What You Should Know About Real Estate Valuation. Investopedia. <https://www.investopedia.com/articles/realestate/12/real-estate-valuation.asp>
- Lifang, Z., Likeya, M., Yang, L., & Dorgizaba. (2023). Residential Rent in Wuhan Based on Geometric Brownian Motion Model Fluctuation Fitting and Prediction. *8th International Conference on Big Data Analytics*, 87–93. Harbin, China: IEEE.

Nei, P. G., Tobi, S. M., & Jasimin, T. H. (2023). Big Data Application in Automated Valuation Model for Valuation Process. *Journal of Advanced Research in Business and Management Studies*, 30(1), 1–15. <https://akademiabaru.com/submit/index.php/arbms/article/view/4684>

Saral, K. (2024). Analyzing the Relationship between Real Estate Investments and Portfolio Diversification. *International Journal of Scientific Research in Engineering and Management*, 8(5), 1–23. <https://doi.org/10.55041/IJSREM32966>

Shbool, M. A., Al-Dmour, R., Al-Shboul, B., Albashabsheh, T. N., & Almasarwah, N. (2024). Real Estate Decision-Making: Precision in Price Prediction Through Advanced Machine Learning Algorithms. *International Journal of Housing Markets and Analysis*. <https://doi.org/10.1108/IJHMA-01-2025-0004>