



Received: 27-04-2025
Accepted: 07-06-2025

ISSN: 2583-049X

Applying Big Data for Developing Real Estate Market in Vietnam: benefits and Challenges

Mai Hoai Dan

University of Finance – Marketing, Ho Chi Minh City, Vietnam

Corresponding Author: **Mai Hoai Dan**

Abstract

This article conducts an in-depth and comprehensive analysis of the impact of big data on the real estate industry and urban development process, especially in the context of increasingly strong digital transformation. The content not only focuses on evaluating basic criteria in data collection and processing but also delves into the specific challenges and limitations that the real estate industry is facing in this process. Issues such as accuracy, reliability, huge data volumes, and security are thoroughly analyzed. Furthermore, the article includes a series of global surveys

on practical applications of Big Data in real estate businesses, aiming to clarify the extent to which big data is being used to drive innovation. And optimize business performance. From there, the article will propose breakthrough solutions, focusing on combining the use of Big Data with advanced technologies such as Artificial Intelligence (AI) and the Internet of Things (IoT). The goal is to create an intelligent ecosystem where data is fully utilized to predict market trends, manage assets more effectively, and enhance customer experience.

Keywords: Big Data, Digital Marketing, Proptech, Real Estate Marketing

1. Introduction

Big data-driven technologies are transforming the real estate industry by offering innovative methods for property development, valuation, sales, and acquisition. The vast volumes of data generated daily are reshaping business strategies across a range of industries, including finance and travel, and are now beginning to make significant strides within the real estate sector, a multi-billion-dollar industry that spans diverse markets. As John D'Angelo, leader of Deloitte's real estate practice in New York, notes, the real estate sector has historically lagged in embracing cutting-edge tools and technologies. However, this is rapidly changing, with both established technology firms and large real estate enterprises increasingly turning to artificial intelligence (AI) and big data analytics to gain valuable insights. While the understanding and application of big data have been somewhat delayed compared to conventional data analytics, this gap is narrowing, and the rise of big data has quickly captured the attention of both academia and industry. Its potential for revolutionizing real estate practices has spurred a surge in both theoretical research and practical implementation. Despite the relatively limited body of academic work focused on big data applications in real estate, the industry is actively seizing this opportunity to drive transformation. The continued growth and refinement of big data technologies in real estate present both exciting prospects for innovation and a critical area for future academic investigation.

McKinsey was among the first to introduce the concept of big data, recognizing its growing importance as a key factor of production that permeates various industries and sectors. Today, big data is a transformative force, reshaping business models and unlocking new opportunities for productivity growth and consumer surplus (Manyika *et al.*, 2011) [6]. Viktor Mayer-Schönberger, a prominent scholar in the field, has become known as a leading figure in the big data era. He has argued that prediction lies at the core of big data, advocating for a shift in focus from pattern recognition to aggregated data, moving away from the pursuit of accuracy toward embracing uncertainty and emphasizing correlation over causality. These conceptual shifts are essential for navigating the complexities and vastness of big data. Mayer-Schönberger has also explored how businesses can harness the power of big data to make informed decisions that can positively influence the future. In 2008, *Nature* published a special issue examining the potential and challenges of big data, addressing its implications in various fields, including Internet technology, biomedical sciences, environmental studies, and cloud computing. *Science*, another leading

academic journal, similarly devoted a special issue to the emerging big data revolution. In 2012, the European Academy of Mathematics and Informatics systematically studied big data, covering topics such as management strategies, academic research trends, and potential outcomes (Li & Xueqi, 2012) [5]. On the practical front, Lohr (2012) argued that big data can provide far more accurate forecasts of real estate market trends compared to traditional economic models, suggesting that big data could disrupt conventional ideas of data ownership. The Ministry of Industry and Information Technology (MIIT) of China, in its "Twelfth Five-Year Plan" of 2012, emphasized key technological innovations, such as advanced data storage and processing technologies, which are integral to the rise of big data. As we move further into the cloud era, big data has shifted from a technical buzzword to a central focus of academic and industry research. In recent years, domestic research has centered on areas with abundant data availability, such as library management systems, digital library construction, the impact of big data on culture and communication, business marketing, micro-credit management, and public health. While progress has been made, the application of big data varies across industries, with certain fields making more substantial advancements. For instance, while many portal sites now support big data sharing, academic research on the valuation and pricing of big data remains limited. Similarly, research on data-sharing scopes or transaction models is still underdeveloped. In the realm of real estate, industry professionals are increasingly leveraging big data in their operations, yet academic research on the subject has not kept pace with the rapid adoption of big data by real estate companies. This gap underscores the need for further exploration and systematic study of the role big data can play in real estate, both from a practical and theoretical standpoint.

2. Research Methods

This study primarily relies on secondary data sourced from credible publications and studies that examine the outcomes of policy implementation. The research methodologies adopted include statistical analysis, comparative analysis, as well as data synthesis and interpretation. The core aim of this study is to develop a methodology capable of effectively addressing the research questions framed around the identified problem statement. In pursuit of this objective, this section details the methodological framework employed. The study design encompasses several key components: The processes of data collection, data analysis, and the assessment of the reliability and validity of the chosen research methods. Each of these elements has been carefully crafted to ensure the research is both robust and credible. This section offers a structured approach to the research, aimed at guiding the analytical process and enabling the derivation of meaningful, evidence-based conclusions.

3. Literature Review

Recent technological advancements, coupled with enhanced data collection techniques, have brought about significant transformations in both the theoretical and applied aspects of real estate evaluation. Pérez-Rave *et al.* (2019) [9] investigated the use of machine learning for big data regression analysis in predicting real estate prices. Their study underscores the increasing volume of data being

collected, processed, and analyzed to forecast market trends and assess consumer sentiment. The integration of big data has the potential to reshape decision-making processes within the real estate sector. However, big data varies substantially in scale and categorization across industries, and broader discussions often overlook critical factors relevant to property analysis. This highlights the need for a comprehensive understanding of the big data it encompasses, how it is applied in real estate, and the ways it is transforming the market. Wei *et al.* (2022) [10] examined the role of price-based models in real estate appraisals in the age of big data. A key point in their research is the inherent uncertainty associated with big data, particularly in terms of its accuracy and reliability. Organizations operating within competitive markets must recognize that simply accumulating big data does not guarantee favorable business outcomes. This suggests that contemporary institutions should move beyond the sole processing of existing data to enhance their financial and economic analyses. Instead, they should focus on fully leveraging available data for segmentation, constructing multiple models tailored to smaller entities, and using advanced information technologies to tackle specific challenges. By doing so, these organizations can better harness the transformative potential of big data, enabling more informed decision-making and improving overall business performance.

Meir *et al.* (2019) [7] investigated the role of data mining in evaluating real estate investment strategies, recognizing that big data is continually produced through real-time data streams and complex event processing. The study identifies two fundamental challenges associated with these data streams: The necessity of real-time data analysis for immediate application and the need for batching and storage of data for subsequent analysis. In this context, big data is distinguished by its variety, encompassing both structured and unstructured data. While structured data adheres to predefined formats and schemas, unstructured data does not conform to any such standards, presenting additional complexities in its processing. Furthermore, the veracity of big data, with its inherent uncertainty, poses a significant concern. Effective data-cleaning processes are essential to ensure that the data remains relevant and reliable for the purposes it is intended to serve. The authors highlight that, despite these challenges, big data and its associated methodologies continue to evolve, with growing recognition of their potential across industries. The application of big data strategies in management, as discussed by Meir *et al.* (2019) [7], involves specialized analytics that differs significantly from traditional data analysis techniques. These advanced analytics enable organizations to derive insights that can offer a competitive edge, particularly when the approach is meticulously structured. By capitalizing on the capabilities of big data analytics, managers can enhance organizational agility, enabling more informed, timely decision-making in a rapidly evolving business environment.

3.1 Benefits of Big Data in Real Estate

The rapid advancements in data analytics have prompted a significant shift in the real estate sector, driving the need for improved efficiency in property appraisals and continuous market monitoring. Kang *et al.* (2021) [3] explored the impact of machine learning and big data on housing price trends, emphasizing that data-centric strategies are crucial for expanding real estate market reach. These strategies

enable more accurate property valuations by leveraging data gathered or shared across different stakeholders. As a result, property managers and investors gain better control over their portfolios and enhance their client interactions. In an industry as expansive as real estate, data plays an essential role in decision-making. Key metrics, such as property valuations, property availability, ownership rates, and other relevant indicators, provide valuable insights. Wei *et al.* (2022) ^[10] note that having access to this type of data significantly enhances business decision-making. However, a major hurdle remains: Obtaining accurate, up-to-date data is a challenging task. Despite the recognized potential of big data analytics to optimize portfolio performance, real estate investors often struggle to acquire the high-quality, current data required for informed decision-making.

Cheryshenko and Pomernyuk (2021) ^[2] explored the role of big data in reshaping decision-making processes within the real estate sector. They highlight that the rapid expansion of data accessibility, coupled with the ability to collect and measure diverse types of information, offers unprecedented opportunities for innovation and the creation of novel revenue models in the real estate market. Currently, real estate developers, agencies, and property managers are already leveraging big data to expand their operations, improve business models, and enhance strategic growth. Big data facilitates more rational, tailored asset management, leading to heightened satisfaction among both investors and clients. Furthermore, real estate firms are able to diversify their investment portfolios and gain deeper insights into individual consumer preferences, fundamentally transforming traditional business practices. These advanced tools also foster a shift in mindset, unlocking new revenue streams for real estate entities. Nevertheless, it is essential to approach the use of big data and analytics with caution, particularly when considering the relevance and quality of the available data across various applications. The selected data must provide real value to all stakeholders in the real estate sector, ensuring that it enhances decision-making and contributes to strategic advancement.

Big data serves as a crucial enabler of strategic development in the real estate market, offering significant advantages to industry stakeholders. It empowers real estate entities to diversify their investments and extract potential value through data mining techniques. As Pai and Wang (2020) ^[8] note, the use of digital personal data and the transformative shifts in business thinking have unlocked new avenues for revenue generation, presenting realty firms with unprecedented opportunities in the era of big data. However, the complexity of managing digital data has grown exponentially in recent years. Winson, Geideman, and Krause (2016) ^[11] highlight that traditional statistical approaches used in real estate have limitations when it comes to processing large and diverse data sets. By integrating advanced big data tools with cutting-edge technologies, real estate firms can efficiently gather, consolidate, and analyze extensive data, thereby unlocking insights that were previously out of reach. In this regard, data analysis has become essential for identifying optimal opportunities while mitigating risk. The vast volumes of historical and digital data collected by real estate professionals can be leveraged through big data and artificial intelligence to generate substantial benefits, enabling more informed decision-making and driving success in an increasingly competitive market.

Big data is revolutionizing the real estate market by enhancing operational efficiency and reducing the risk of poor investment decisions. It enables firms to offer clients more accurate and profitable property recommendations, minimizing the likelihood of selecting suboptimal assets. Through predictive analytics and sentiment analysis, big data allows for the delivery of personalized property options, providing real estate companies with a competitive edge in a crowded market. The inherent volatility of the real estate market, often marked by fluctuating property prices, can be mitigated through big data's ability to forecast future trends. By analyzing real-time data, firms can make more informed decisions, conduct region-specific price evaluations, and predict market shifts, enabling them to offer tailored solutions like personalized insurance coverage with future property value projections. To maximize client engagement and comprehension, real-time insights must be presented in a clear and impactful manner. This is why leading companies like Zillow are integrating AI-driven big data tools to optimize customer engagement and generate higher profits. Additionally, big data opens new avenues for revenue optimization within the real estate sector. The primary goal of any organization is to maximize return on investment, and big data offers a powerful tool to drive growth across all facets of real estate operations. By utilizing advanced analytics, companies can reduce investment risk, enhance profitability, and ensure long-term sustainability (Barkham *et al.*, 2022). The rise of digital intelligence marks a paradigm shift in real estate business strategies. As big data becomes more integrated into decision-making processes, it is reshaping the management of risk, responsibility, and client needs. This transformation is accelerating innovation and improving service delivery. Notably, international real estate firms have embraced big data not only to meet client demands but also to enhance transaction efficiency. By focusing on customer satisfaction and leveraging data-driven insights, these companies are setting new standards for success in the industry.

3.2 Challenges in using Big Data in Real Estate

The process of data collection, organization, and analysis constitutes a critical phase in unlocking its potential for knowledge extraction. Modern technological advancements, including both software solutions and cloud-based platforms, have become integral in facilitating this process within the real estate sector (Kok *et al.*, 2017) ^[4]. These tools provide capabilities for data visualization, filtration, analysis, and simulation of future events, enabling stakeholders to assess market trends and financial properties, and design decisions with enhanced precision. By leveraging these technologies, industry professionals can make informed predictions about potential future outcomes, which is poised to significantly improve decision-making practices. However, Kok *et al.* (2017) ^[4] point out the intense competition among various technological solutions aiming to dominate the real estate market. Traditional software systems are increasingly being supplanted by cloud-based platforms, which offer enhanced capabilities for automating the aggregation of data into large, valuable datasets. Although big data forecasting holds considerable promise as a game-changer within the real estate market, various entities within the industry are confronted with challenges related to reputation and trust. A key issue is the persistence of conventional risk analysis models, which

remain the standard approach within many segments of the industry. Another significant challenge relates to the integrity and reliability of the datasets being used for decision-making. Rossini (2000) discussed the application of expert systems and artificial intelligence in the realm of real estate predictions, emphasizing the complexities associated with aggregating large volumes of data over extended periods. One of the central difficulties lies in the assessment of the quality of such datasets. While data filtering mechanisms can be beneficial, they are not without their limitations, especially in terms of ensuring data accuracy and completeness. Moreover, the precision of forecasting models is often contingent on the quantity of data being processed. As such, the ability to maintain forecasting accuracy becomes increasingly difficult as datasets grow in scale and complexity. Perhaps one of the most formidable challenges is the development of robust mechanisms to cope with the continual expansion of datasets. As the volume of data continues to increase at an exponential rate, the real estate industry faces significant hurdles in managing, processing, and forecasting this data effectively. To capitalize on the full potential of big data, it is imperative that the industry adopts advanced technological solutions, reevaluates traditional risk models, and works to enhance the integrity and quality of the data being utilized for critical decision-making. In summary, while big data holds considerable transformative potential for the real estate industry, several challenges must be addressed to realize its full benefits. These include the competition among technological solutions, concerns regarding data integrity, and the growing difficulty of managing and processing increasingly large datasets. Overcoming these obstacles will require a comprehensive approach that incorporates both technological innovation and the improvement of data management practices. By doing so, the real estate industry can harness the power of big data to drive more accurate, reliable, and informed decision-making processes.

The primary challenge in the effective application of big data within the real estate market is the interpretation of the data itself. As highlighted by Ali and Siniak (2020)^[1], there is a growing trend in the use of large-scale urban data and its influence on real estate markets. The authors emphasize that the interpretation of big data must be conducted with intelligence, and security, and from a perspective of aggregation, ensuring that the complexities of the data are comprehensively addressed. A critical issue that arises from this process is the difficulty in determining the appropriate analytical methods and deriving actionable insights from the datasets. Many organizations refrain from utilizing big data analytics due to the substantial investment required for the supporting infrastructure and the complexity involved in its implementation. Furthermore, an unresolved challenge lies in the inherent contradictions between privacy protections and the use of big data. The processing technologies required to handle big data are not readily accessible to all real estate firms, particularly smaller entities. Additionally, the unique characteristics of real estate data, including its volume and complexity, add another layer of difficulty in its utilization. Big data in real estate encompasses far more than merely large quantities of numerical data; it is often unstructured and complex, requiring advanced technologies for its collection, storage, and processing. These demands present a considerable challenge for real estate corporations

attempting to extract meaningful insights from vast datasets. Given the complexity and scale of big data, it may be more pragmatic for real estate firms to entrust the management and processing of such data to specialized third-party providers who possess the requisite technical expertise and infrastructure. As noted by Wei *et al.* (2022)^[10], different types of entities play distinct roles in the big data ecosystem, and real estate firms can benefit from focusing on the application of insights derived from data rather than attempting to manage the intricate task of data processing in-house. By delegating the technical aspects of big data processing to technology vendors, real estate organizations can concentrate on leveraging the insights to enhance business strategies, drive development, and improve decision-making.

Coleman *et al.* (2022) examined the integration of machine learning techniques for assessing real estate prices through the application of location-based big data. Their research underscores that, while the real estate sector is a pivotal component of the real economy, it exhibits distinctive characteristics commonly associated with fictitious economies, such as complexity, elevated risk, and metastability. These features introduce substantial uncertainty in the application of big data, as metastable systems are highly sensitive to shifts in anticipations and expectations. The dissemination and public sharing of big data can, therefore, significantly influence investment demand, leading to alterations in market behavior. Such changes can destabilize the real estate market, thereby posing potential risks to broader economic stability. Moreover, Coleman *et al.* (2022) identify critical challenges inherent in the structural imbalances and information asymmetry that emerge when managing big data in the real estate sector. They advocate for the establishment of a dedicated platform for the storage and aggregation of real estate data, which would facilitate the creation of comprehensive big data repositories. The study also highlights the growing trend within the real estate market towards enhanced business management and collective data integration, which positions big data as a central element for operational efficacy. Real estate firms are thus encouraged to develop data warehouses and adopt integration management practices, utilizing advanced data mining and analytical techniques to generate more accurate market forecasts. In the context of property valuation and appraisal, various big data models are employed within the real estate industry, with Automated Valuation Models (AVMs) being the most widely utilized. As articulated by Kok *et al.* (2017)^[4], an AVM is a service that applies mathematical models to estimate the value of a property or asset at a specific moment in time. AVMs may adopt diverse methodologies for valuation, with regression and hedonic models being the predominant approaches. More recently, some AVMs have integrated adaptive estimation models or neural, self-learning network systems to enhance model precision and adaptability (Kok *et al.*, 2017)^[4]. Hedonic AVMs, in particular, are extensively recognized in the real estate sector. These models rely on detailed information regarding the intrinsic characteristics of properties, such as the year of construction, size, and extent of capital improvements. Hedonic AVMs typically employ search engines to compare these attributes with those of similar properties within a defined radius, facilitating a comparative analysis. One of the key advantages of hedonic AVMs is their simplicity and

ease of implementation, as they are grounded in regression models that are conceptually straightforward and computationally efficient. However, Kok *et al.* (2017) [4] also highlight significant limitations associated with the use of regression-based AVMs. These models are inherently universal, applying a single predictive formula across all variables. While this characteristic enhances the general applicability of the model, it may lead to inaccuracies, particularly in cases where the property in question deviates from the assumptions underlying the model. Moreover, the use of a single model across diverse property types and market conditions may not capture the subtleties that are crucial for providing accurate property valuations.

The application of machine learning in the real estate market for big data analysis has emerged as a highly promising approach due to its substantial predictive capabilities. As Pérez-Rave *et al.* (2019) [9] note, machine learning techniques typically utilize two distinct subsamples: One designated for training the model and the other for validation purposes. The integration of hedonic regression with machine learning algorithms offers significant opportunities for improving the analysis of real estate prices. This integration facilitates a more robust synthesis of both inferential and predictive analytics, leveraging the models that are developed. Nevertheless, this integration process is not without its complexities. A key challenge arises from the limitations of inferential analysis when applying machine learning-based regression to big data. Specifically, the coefficient estimates derived from machine learning models often exhibit inconsistency, failing to satisfy essential statistical prerequisites for reliable inferential analysis. Moreover, regularization methods designed to reduce overfitting in non-training data samples can induce bias by excluding important variables from the model. This issue is further compounded by the high correlations between variables, which leads to instability in the resulting models. As a consequence, the selected predictors may vary significantly across different samples, leading to the development of models that are not consistent. Kok *et al.* (2017) [4] assert that these shortcomings underscore the necessity of incorporating additional strategies to enhance decision-making processes and further the advancement of knowledge within this field. To address the limitations inherent in the combination of hedonic regression and machine learning for big data analysis, one pivotal solution lies in the strategic selection of variables. This process involves identifying a subset of variables that are deemed most crucial for comprehensively understanding the underlying phenomenon. Such a variable selection process is integral to improving the precision and reliability of both inferential and predictive models, ensuring that the variables incorporated into the analysis are directly pertinent to the real estate market dynamics.

3.3 Trends in Adoption of Big Data in Real Estate

A thorough review of the available literature and emerging trends regarding the application of big data in the real estate industry reveals that while there are numerous theoretical propositions and speculative analyses, there is a notable lack of empirical evidence confirming the widespread implementation of big data solutions within the industry. Although certain real estate entities may have initiated the adoption of big data technologies, there is little public disclosure of such efforts, leaving the true extent of their

adoption largely unknown. Given this, it is plausible to conclude that, on the whole, the real estate sector has not yet fully integrated big data into its core operations.

Despite the absence of concrete data regarding current industry practices, a significant body of commentary and analysis has emerged, discussing the substantial potential that big data could unlock for the real estate sector. Key areas where big data is anticipated to make a meaningful impact include the enhancement of property valuation models, more sophisticated and targeted property marketing strategies, and improvements in the pricing mechanisms for property insurance (Cheryshenko & Pomernyuk, 2021) [2]. The integration of big data could also facilitate the development of innovative real estate products, tailored to meet the specific needs of individual consumers, alongside offering advanced quality management systems for property maintenance and improvement. Moreover, big data could play a crucial role in optimizing occupancy rates and maximizing rental yields, as it would provide a more granular understanding of market dynamics, consumer preferences, and regional trends. However, it is essential to recognize that much of the discussion surrounding these potential benefits remains speculative, with no substantial evidence to suggest that such outcomes have been realized on a large scale within the industry.

The complexity and resource intensity associated with the implementation of big data solutions further complicate the widespread adoption of these technologies. The development and deployment of effective algorithms for big data applications require substantial time, expertise, and financial investment. Similarly, data mining, a fundamental component of any big data solution, can be prohibitively costly and resource-intensive. Despite these challenges, many real estate businesses currently utilize traditional analytics methods to project demand trends, price fluctuations, and market conditions. For a significant portion of the industry, these existing tools may already provide sufficient value, reducing the perceived need for further investment in big data infrastructure. Consequently, there is a discernible reluctance among real estate professionals to make the substantial financial and operational commitments required to integrate big data solutions into their decision-making processes.

The findings from the analysis suggest that the real estate industry's engagement with big data remains in its infancy. While there is considerable potential for big data to revolutionize aspects of the sector, including property valuation, marketing, and risk management, the industry has not yet fully explored or capitalized on these opportunities. As a result, there exists a substantial gap between the theoretical possibilities that big data offers and its practical implementation within the sector. This gap underscores an important avenue for future research and development. The real estate industry is at a critical juncture, where the integration of big data could lead to significant advancements in operational efficiency, decision-making accuracy, and market responsiveness. Therefore, it remains an opportune moment for the sector to critically evaluate the role of big data and actively pursue its application to drive improvements across a range of operational functions. In conclusion, while the real estate industry has yet to fully embrace big data, the potential benefits that could be realized from its adoption are clear. By harnessing the power of big data, real estate professionals could gain

deeper insights into market trends, consumer behavior, and property valuations, ultimately leading to more informed decision-making and enhanced industry practices. However, for these benefits to be realized, the industry must overcome the challenges associated with big data implementation, including the development of appropriate algorithms, the acquisition of relevant data, and the allocation of resources toward technological infrastructure. Given the promising potential of big data, it is essential that the real estate industry continues to explore and invest in these technologies to maintain competitive advantage and drive future growth.

4. Data Applications in Real Estate Market

Content analysis can be categorized into two main levels, depending on the specificity of information and the depth of analysis required. The first level is thematic analysis, while the second level is keyword analysis. The thematic analysis primarily aims to identify the main themes within the content, whereas keyword analysis isolates key information within each theme and may be used to identify subtle differences in the information presented. This study will focus on thematic analysis (Kuckartz, 2019), as the research aims to identify the key thematic areas under each research question. Content analysis can also be classified into two primary types: Conceptual analysis and relational analysis. Conceptual analysis focuses on identifying the occurrence of the main concept under study, which in this case is the application of big data in real estate. The concept of big data encompasses both machine learning and artificial intelligence, which play a significant role in improving decision-making in real estate valuation and transactions. Therefore, in this study, the concept of big data is presented in the form of a decision model or decision tree, which tracks the information considered by a customer from the point of intention to purchase real estate to the final decision of whether to buy or not buy the property. The decision-making process in real estate is a data-driven journey, in which potential buyers collect and evaluate information from various sources. Typically, this journey begins online, where buyers search for property information, compare values, and read reviews. They then seek additional insights from real estate agents and consult authorities responsible for maintaining property records. By exploring this concept, the analysis will examine the types of information influencing real estate decision-making and how this information can be modeled to support more accurate forecasting and valuation.

Another approach to content analysis is relational analysis, which focuses on examining the relationships between concepts. For instance, when analyzing big data, it is essential to consider related concepts such as decision modeling, machine learning, and artificial intelligence. In real estate decision-making, it is also critical to include factors from areas such as finance, employment, and income levels, as these elements play a significant role in shaping consumer decisions and the strategies used by real estate agents to market properties to consumers. In this study, relational analysis will incorporate cognitive mapping and proximity analysis. The implementation of the content analysis methods outlined above will follow a structured process. The first step is to determine the number of concepts to code for, which will be guided by the research questions. The next step is to identify the themes and reduce

the content to these thematic areas. The third step involves establishing the coding rules, focusing on whether the data should be coded based on the occurrence or frequency of themes. The fourth step is to identify patterns within the themes, and the final step is to explore the relationships between the themes. By applying this analytical framework, it becomes possible to draw inferences about the research subject in a repeatable and verifiable manner.

Machine Learning (ML) is a subset of Artificial Intelligence (AI) that empowers systems to learn from data, adapt, and improve their performance over time without the need for explicit programming. Rather than relying on pre-programmed rules, ML models derive patterns and insights from the data they are exposed to, making them particularly well-suited for handling complex and large-scale datasets. In the context of the real estate industry, the volume and complexity of data, such as property values, market trends, and consumer behavior, make ML an invaluable tool. It facilitates the processing of big data, enabling industry professionals to derive actionable insights and improve decision-making processes. One of the key advantages of ML in real estate is its ability to process large volumes of unstructured and heterogeneous data, which often includes incomplete, noisy, or inconsistent information. Traditional analytical methods struggle to handle such diverse data, but ML can learn patterns and make predictions despite these challenges. Additionally, ML can simplify complex tasks by eliminating the need for a predefined mathematical model, allowing for more flexible and adaptive solutions in real estate applications, such as predictive pricing, demand forecasting, and risk assessment. The field of ML includes several advanced algorithms, each with its own strengths and applications. Four of the most widely used and modern machine learning methods are Artificial Neural Networks (ANN), Support Vector Machines (SVM), Deep Learning (DL), and Ensemble Learning (EL) (Wei *et al.*, 2022) [10]. Each of these techniques has unique characteristics that make them suitable for specific real estate applications.

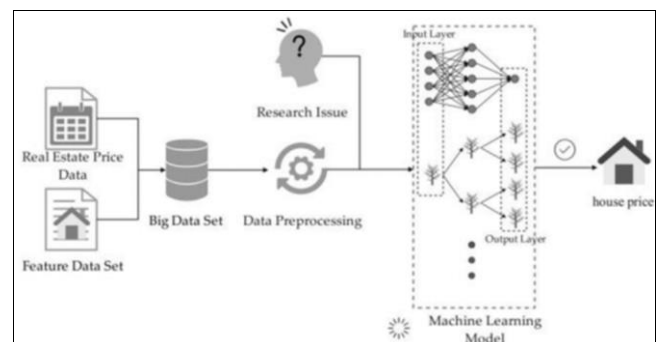


Fig 1: Flow Chart of Applying Machine Learning

The Artificial Neural Network (ANN) is an advanced multivariate analytical tool that has become increasingly prevalent in the real estate industry due to its ability to model nonlinear relationships and detect complex data patterns. ANN operates by mimicking the structure and learning mechanisms of biological neural networks, consisting of input, hidden, and output layers. In the real estate sector, ANN is widely applied in property valuation, market trend analysis, and assessing factors influencing real estate prices. One of its primary advantages is its capability to learn from historical data, thereby enhancing the accuracy

of predictions. However, ANN also presents several limitations, including:

- Support Vector Machine (SVM) is a supervised learning algorithm highly effective in classification and prediction tasks. A key feature of SVM is its ability to optimize learning from small datasets by using a limited number of support vectors rather than the entire data space. This approach helps mitigate the "curse of dimensionality," a common issue when handling real estate data with numerous variables. SVM has been successfully applied in various real estate studies, including property price forecasting, market trend analysis, and risk assessment in real estate investment. For example, research on the Shanghai real estate market utilized SVM to predict housing price fluctuations with higher accuracy than traditional statistical models.
- Deep Learning (DL) is an extension of ANN, incorporating multiple hidden layers to automatically extract features from data without requiring manual preprocessing. This method can process large volumes of data and detect intricate hidden patterns, significantly improving prediction accuracy compared to traditional algorithms. In the real estate sector, DL has promising applications in: (1) Property price forecasting: Leveraging market data, property images, and macroeconomic indicators; (2) Customer demand analysis, using deep learning algorithms to recommend suitable real estate products; (3) Integration of unstructured data, including textual descriptions (property listings), images (property photos), and real-time market data.
- Ensemble Learning (EL) is not a standalone machine learning algorithm but rather a method that combines multiple models to enhance prediction accuracy and generalization capabilities. EL can significantly improve real estate forecasting performance by leveraging the strengths of various algorithms. The three primary strategies in ensemble learning include: (1) Bagging: This technique trains multiple independent models on different subsets of data and then aggregates their predictions. Random Forest is a well-known example of this method, frequently applied in real estate price forecasting; (2) Boosting: This approach builds models sequentially, where each new model corrects the errors of the previous one. Algorithms such as AdaBoost and Gradient Boosting Machines (GBM) have been successfully employed in real estate market analysis. (3) Stacking: This technique combines multiple prediction models and employs a higher-level machine learning model to optimize the final output.

Machine learning (ML) introduces various advanced methodologies to improve the accuracy of real estate price predictions and market analysis. Each method has its unique advantages and limitations, depending on data characteristics and application objectives. In the future, integrating advanced algorithms such as DL and EL could revolutionize the real estate industry, optimizing decision-making processes and mitigating investment risks.

Currently, the real estate industry relies heavily on housing price indices, which are used to track the price movements of land and buildings in different regions. These indices are updated at various frequencies, including monthly,

quarterly, semi-annually, and annually. They are mainly provided by the real estate industry, based on data from recent transactions of land, housing, and commercial real estate. Some indices also track rental and lease prices. However, in most cases, these indices only serve as a reflection of market prices and thereby affect real estate values. The use of housing price indices as the primary data source in the industry has many limitations. In fact, these indices are not designed to support decision-making in the real estate industry but are primarily designed as a tool to assess the health and stability of the market. As a result, some industry players do not have the tools they need to make strategic decisions. Furthermore, the existing data structures in the market tend to favor real estate agents over owners, sellers, or buyers. These shortcomings call for a more comprehensive data system to support decision-making in the industry. A viable alternative is the application of big data to the real estate sector. The concept of big data allows for a broader range of inputs into the real estate valuation process, focusing on factors that buyers and sellers consider when making transaction decisions. The real estate buying and selling process can be viewed as a journey that generates and consumes large amounts of data. The application of big data not only leverages information that users are aware of but also exploits data that they have not yet accessed or considered. This is done through the application of supervised and unsupervised machine learning models, to optimize decision-making in the real estate industry.

5. Conclusions

The study concludes that the level of adoption of big data solutions in the real estate industry remains low, mainly because businesses do not know how to translate big data into specific business objectives. Furthermore, the complexity of data analysis algorithms further increases the barrier to adoption. Implementing big data requires significant investment in highly skilled human resources to develop and train models, in addition to capital investment in technology and data processing time. However, if businesses do not clearly identify the link between data and business results, this investment is difficult to justify. Despite the challenges, big data offers significant potential in the field of risk management. Organizations with the highest risk levels, including mortgage companies, commercial banks, investment banks, and financial institutions involved in real estate securitization, will benefit the most from the application of this technology. In particular, the growth of mortgage-backed securities (MBS) and collateralized debt obligations (CDOs) can amplify financial risks, making the application of big data an important tool to improve the risk assessment and management capabilities of financial institutions.

6. References

1. Ali B, Siniak N. Integrating big data into decision-making in real estate industry. *Management*. 2020; 36(2):16487974.
2. Cheryshenko MS, Pomernyuk YY. Integration of big data in the decision-making process in the real estate sector. In *IOP Conference Series: Earth and Environmental Science* (Vol. 751, No. 1, p. 012096). IOP Publishing, April, 2021.
3. Kang Y, Zhang F, Peng W, Gao S, Rao J, Duarte F, *et*

- al.* Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*. 2021; 111:104919.
4. Kok N, Koponen EL, Martínez-Barbosa CA. Big data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management*. 2017; 43(6):202-211.
 5. Li G, Xueqi C. Research status and scientific thinking of big data. *Bulletin of Chinese Academy of Sciences*. 2012; 27(6):647-657.
 6. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, *et al.* Big data: The next frontier for innovation, competition, and productivity, 2011. Retrieved October 13, 2019, from: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation/>
 7. Meier Z, Pan WT, Zhuohong S, Yingying Z, Zuchang Z. Application of data mining technology in evaluating real estate investment plan based on GRA-AHP. In *Journal of Physics: Conference Series* (Vol. 1284, No. 1, p. 012037). IOP Publishing, August, 2019.
 8. Pai PF, Wang WC. Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences*. 2020; 10(17):5832.
 9. Pérez-Rave JI, Correa-Morales JC, González-Echavarría F. A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes. *Journal of Property Research*. 2019; 36(1):59-96.
 10. Wei C, Fu M, Wang L, Yang H, Tang F, Xiong Y. The research development of hedonic price model-based real estate appraisal in the era of big data. *Land*. 2022; 11(3):334.
 11. Winson-Geideman K, Krause A. Transformations in real estate research: The big data revolution. In *Proceedings of the 22nd Annual Pacific-Rim Real Estate Society Conference, Queensland, Australia, January, 2016, 17-20.*