Audit Financiar, XXIII, Nr. 1(177)/**2025**, *180-196*
ISSN on-line: 1844-8801

Reforming Real Estate Valuation for Financial Auditors with AI:

An In-Depth Exploration of Current Methods and Future Directions

Silviu-Ionuţ BĂBŢAN, Ph. D. Student, Accounting and Audit Department, The Faculty of Economics and Business Administration, Babeş-Bolyai University, Cluj-Napoca, Romania,
e-mail: silviu.babtan@econ.ubbcluj.ro

Abstract

Artificial Intelligence (AI) is changing real estate valuation with innovative approaches. This article examines several AI methods – Regression Models, Decision Trees, Random Forests, Artificial Neural Networks, and XGBoost – and explores their applications for improving property valuation accuracy and efficiency, with implications for other professions involved, e.g. audit. The author starts by investigating traditional valuation methods' limitations, such as data constraints and subjectivity, and presents how these AI techniques, which are translated in property valuation field as automated valuation methods, tackle these challenges. Regression Models quantify attributes, Decision Trees provide clear insights, Random Forests improve predictions, Artificial Neural Networks design elaborate relationships, and XGBoost furnishes advanced boosting techniques for higher performance. Underscoring that AI is meant to support, not substitute, human assessors, the paper presents how these methods can enhance valuation processes, deliver more reliable valuation reports, and decrease errors, while also exploring future innovations and evolving trends in artificial intelligence for real estate industry and related professions.

Key words: *artificial intelligence; real estate valuation; audit, automated valuation techniques methods;*

JEL Classification: *R30, C40, M40*

To cite this article:
Băbţan, S.-I. (2025), Reforming Real Estate Valuation for Financial Auditors With AI: An In-Depth Exploration of Current Methods and Future Directions, *Audit Financiar*, vol. XXIII, no. 1(177)/2025, pp. 180-196,
DOI: 10.20869/AUDITF/2025/177/005

To link this article:
http://dx.doi.org/10.20869/AUDITF/2025/177/005
Received: 11.09.2024
Revised: 30.09.2024
Accepted: 20.01.2025

Introduction

Artificial intelligence (AI) is a fast-transforming field that is making notable influences across various areas, fundamentally changing how activities are accomplished not only in business, but in the economy (Svetlana et al., 2022).

In this research, we will explore how various AI methods can be used for real estate valuation. The real estate field is dynamic and complex, with property values influenced by a variety of factors such as property size, age of the building, location, economic conditions, and market trends. Traditional valuation methods depend on expert judgment and manual assessments, which may be inconsistent, time-consuming, and subject to human error (Choudhury, 2015). Within this framework, there is an intensifying necessity for applying automated methods in real estate valuation (AVM). These methods, machine learning models and advanced algorithms, can examine vast volumes of data in a short time and accurately, offering objective and consistent valuations (Zhou et al., 2017).

AVM boost efficiency, improves the reliability of property assessments, and decreases costs. Therefore, it is a precious tool for real estate stakeholders and professionals in evaluating investment options and pricing strategy. Secondly, when auditing the financial reports of businesses that include real estate properties, auditors offer particular focus to the property valuations for these assets. If the market value of real estate is shown in the financial reports along with historical cost, auditors must assure us that these valuations are precisely determined and represent market conditions. For this purpose, auditors analyze the appraisal reports to evaluate the methodologies used by independent appraisers, as well as the expectations and data inputs applied in estimating the fair market value. They focus on factors such as the similarity of market data, the suitability of the appraisal approach (e.g., income, market comparison, or cost methods), and whether these methods have been applied precisely in accordance with valuation standards (Brown, 2019)

Also, auditors evaluate if appraisers have examined all relevant aspects that could impact market value, such as up-to-date economic conditions, specific local market trends, and the property's unique attributes. To validate the estimates in the valuation reports, auditors could compare the results with other similar appraisals or consult independent professionals. They also assure that the valuation reports are detailed sufficiently and that the financial report disclosures clearly outline how the market value was established, including any variables or potential changes This meticulous verification is fundamental to confirm that the values reported in the financial reports align with market conditions and to reduce the risk of inaccurate audit reporting, which could influence the decisions of shareholders and other stakeholders (Choudhury, 2015).

Despite these benefits for property valuation and audit professions, the effectiveness of AVM is dependent on the quality of the database and the technical expertise of the individuals who implement these methods. By investigating these AI-driven techniques, our main aim is to identify AVM that can improve the precision and the performance of the real estate valuation process, contributing to more comprehensive and reliable valuation reports (Zhang, 2018).

The paper aims to provide significant insights into how AI can revolutionize real estate valuation processes with a significant impact on accounting and audit professions which verify the fair value estimations. As we delve into the intricacies of Artificial Intelligence implementation in property valuation, one question stands out: What are the automatic methods that can be used in the valuation process? To address this question, we analyzed complex statistical methods presented in specialized literature that were used for estimations in other fields of activity, such as finance, trade or the capital market.

Regarding the research methodology, to determine relevant specialized literature on Artificial Intelligence techniques, we conducted a narrative approach using Google Scholar. As recommended by Ferrari R. (2015), in order to increase the performance of the narrative approach, we borrowed elements from the systematic review methodology. Therefore, our researched was performed utilizing the terms: Artificial Intelligence methods, linear regression prediction, decision tree prediction, random forest prediction, artificial neural network prediction, and Extreme Gradient Boosting. In our paper, we included only reviewed journal articles focusing on the mentioned AI techniques. The papers were also required to address benefits and limitations and incorporate performance metrics. Papers without evaluation criteria or not focused on AI for estimation processes were excluded. Also, the studies from grey literature were excluded. This information was used to compare and analyze the methods across application areas, identifying advantages and disadvantages. We also conducted a critical quality assessment to prioritize reviewed papers with a clear and transparent methodology.

The element of novelty brought by our paper is that it brings together all the automatic estimation methods presented in the specialized literature. In addition, it presents the advantages and disadvantages of each presented technique, as well as recommendations regarding the application of the method for value prediction. By analyzing and studying various AI-driven models, the study wants to illustrate that these technologies can be efficient, reliable, and flexible solutions to fulfill the shifting demands of the real estate industry. The result of our analysis consists in establishing a clear working methodology for the application of automatic property valuation methods, regardless of the area in which they are located or the period.

The paper is organized as follows: Section 1, which provides an overview of AI, encompassing conceptual definitions, classifications, and diverse application areas. Section 2 presents the context of our discussion. Section 3 explores the applications of AI in real estate valuation. It analyzes comprehensively each automated valuation model, displaying how they work, their advantages, and their disadvantages. This section's aim is to offer an exhaustive analysis of the practical application and challenges related to different AI techniques in the context of real estate valuation, providing significant insights into their potential limitations and effectiveness. The last section concludes the research by outlining the insights and the key findings. It encompasses the outcome, considering the implications of AI in the process of real estate valuation and recommending future directions of research.

1. Artificial Intelligence (AI): concept explanation, classifications, and fields of application

AI encompasses the examination and development of automated systems and software able to learn, reason, acquire knowledge, manipulate objects, communicate, and perceive their environment (Pannu, 2015). AI is increasingly significant in management science and operations research, where intelligence is usually identified as the capacity to accumulate knowledge and utilize rationality to solve complex issues.

In Table no. 1, the broad field of AI is orderly classified into distinct sub-categories and domains, providing a detailed framework that encapsulates the varied applications and methodologies fundamental to AI. This overview not only underlines the diversity within AI but also supports a clearer comprehension of its complex nature.

Table no. 1. AI areas

Category	Sub-Categories
A. Cognitive Science Applications	• *Learning Systems,* • *Intelligent Agents,* • *Expert Systems,* • *Genetic Algorithms,* • *Neutral Networks,*
B. Natural Interface Applications	• *Natural Languages,* • *Virtual Reality,* • *Speech Recognition,*
C. Speech Understanding & Semantic Processing	• *Language Translation,* • *Speech Understanding,* • *Information Retrieval,* • *Semantic Information Processing,*
D. Learning and Adaptive Systems	• *Concept Formation,* • *Cybernetics,*

Category	Sub-Categories
E. Problem Solving	• *Inference,* • *Automatic Program Writing,* • *Heuristic Search,* • *Interactive Problem Solving,*
F. Perception (Visual)	• *Scene Analysis,* • *Pattern Recognition,*
G. Modeling	• *The Representation Problem for Problem Solving Systems,* • *Modeling Natural Systems,*
H. Robotics Applications	• *Dexterity,* • *Visual Perceptions,* • *Navigation,* • *Locomotion,*
I. Robots	• *Industrial Automation,* • *Exploration,* • *Transportation/Navigation,* • *Military,* • *Security,* • *Household,* • *Other,*
J. Games	• *Games,*

Source: Author's own composition, based on Khanzode et al. (2020) and Pannu (2015)

As illustrated in **Table no. 1**, AI encompasses an extensive variety of areas, ranging from virtual reality and robotics deployment that optimize technical and industrial processes, to the examination of visual data and the generation of forecasting techniques. This variety showcases the wide coverage and diverse applications of AI. The table shows that within different branches of AI, Cognitive Science Applications can be successfully implemented in real estate valuation. This category incorporates fundamental AI techniques that encompass learning models and systems, which are used for a diversity of estimation and predictive tasks.

In accounting and audit, the introduction of AI has generated concerns among experts about potential employment displacement (Mohammad et al., 2020). Nevertheless, a more refined perspective indicates that AI will not supplant assessors and accountants but will enhance their capabilities. AI can manage time-consuming and routine tasks, enabling accountants and auditors to concentrate on more value-added and intricate activities. This change can lead to improved accuracy and performance, reducing the work time spent by accounting professionals, and, in the end, enhancing the overall effectiveness of the accounting industry.

The Artificial Intelligence methods that we proposed in this paper have been successfully implemented in other fields of activity. As a result, we emphasize the possibility of integrating these methods in the property evaluation process. **Table no. 2** highlights the main fields where these methods were successfully applied.

Table no. 2. Practical implementation of AI methods

Author	AI Method	Prediction of
Goundar S. et al. (2021)	*Linear Regression*	*Property Valuation*
Boztosun D. et al. (2016)		*Economic Growth*
Zhou T. et al. (2013)		*Carbon Sink Strength*
Roy S. et al. (2015)		*Stock exchange rates*
Saini D. et al. (2016)		*Electricity Price*
Ge Y. et al. (2020)		*Corn Price*
Khan Z. et al. (2022)		*Used Car Price*
Manoj J. et al. (2019)		*Price of Gold*
Oba K. M. (2019)		*Cement Price*

Author	AI Method	Prediction of
Lasota T. et al. (2013)	*Decision Trees*	*Property Valuation*
Padmanaban K. A. et al. (2016)		*Chronic Kidney Disease*
Ghosh A. et al. (2021)		*Soil Erosion Risk*
Aji N. A. et al. (2019)		*Credit Scoring*
Bhatnagar R. et al. (2020)		*Crop Yield*
Sisodia D. et al. (2018)		*Diabet*
Putra P.H. et al. (2023)		*Car Price*
Vaiz J.S. et al. (2016)		*Stock Price*
Nwulu N.I. et al. (2017)		*Oil Price*
Goundar S. et al. (2021)	*Random Forest*	*Property Valuation*
Langsetmo L. et al. (2023)		*Hip Fracture Risk*
Langsetmo L. et al. (2023)		*Mortality Risk*
Khaidem L. et al. (2016)		*Stock Market Price*
González C. et al. (2016)		*Electricity Price*
Ghosh A. et al. (2021)		*Soil Erosion Risk*
Aji N. A. et al. (2019)		*Credit Scoring*
Bhatnagar R. et al. (2020)		*Crop Yield*
Putra P.H. et al. (2023)		*Car Price*
Shanbehzadeh M. et al. (2022)	*Neural Network*	*Mortality Among Covid-19*
Yan K. et al. (2019)		*Energy Consumption*
Khan Z. H. et al. (2011)		*Price of Share Market*
Jha G. K. et al. (2013)		*Agricultural Price*
Ugurlu U. et al. (2018)		*Electricity Price*
Nikolaev D. et al. (2021)		*Equity Price*
Zhou Y. et al. (2019)	*Extreme Gradient Boosting*	*Crude Oil Price*
Ma B. et al. (2020)		*Diagnostic Classification of Cancers*
Voung P.H. et al. (2022)		*Stock price*
NandigalaVenkatAnurag Y. et al. (2019P)		*Air Quality Index*
Ramani K. et al. (2023)		*Bitcoin Price*

Source: Author's own composition

As demonstrated in **Table no. 2**, AI methods are used in estimation in various fields, from the medical field to the economic, financial, or energy field. We noticed in the specialized literature, that all the methods proposed by us for real estate valuation have already been used in price estimations in other fields, like stock price (Vaiz et al., 2016, Voung et al., 2022), gold price (Mombeini et al., 2015, Manoj et al., 2019), electricity price (Saini et al. 2016, González et al. 2016*)* or even Bitcoin price (Ramani K. et al. 2023). Consequently, we consider that these methods can be practically implemented in the valuation of real estate properties.

There is also a synergy between real estate valuation and audit. To understand the goal of valuation, it is essential to refer to valuation standards and concepts, which offer the conceptual foundations of this method. The main objective of property valuation is to establish its value within a specific context, whether it is for financing, sales transactions, financial reporting or taxation (Smith, 2020). Particularly, in the context of financial reporting, valuation goals to reflect a fair market value that is useful and relevant to the users of financial reports, such as creditors, investors and other stakeholders (Johnson and Williams, 2021).

The roles of the auditor and the appraiser intersect in a crucial way. The appraiser is responsible for using methodologies to determine the fair market value of a property, taking into consideration all relevant market factors, including actual economic conditions and the specific attributes of the property (Brown, 2019). On the other hand, the auditor is tasked with validating and verifying this valuation, ensuring that the used method is accurate and that the results are accurately reflected in the financial reports. Therefore, the collaboration between the auditor and the appraiser is crucial to ensure that the values reported in the financial reports are precise, consistent with market realities, and compliant with financial reporting standards and accounting (Davis and Taylor, 2022).

2. Artificial Intelligence in real estate valuation

In recent decades, AI has begun transform various sectors, including property valuation. The application of AI in this area offers significant advantages, such as increased efficiency and accuracy in determining the fair value of real estate properties. This is important for financial auditors who want to validate the correct value assessments shared in companies' financial reports (Smith, 2020).

AI permits the automated valuation of properties by applying complex machine learning algorithms that examine current data and historical on real estate transactions. These algorithms could rapidly process massive amounts of data, delivering relatively and quick precise estimates. For example, by examining data on location, sale prices, features, size and property condition, AI can generate market value estimates used by both auditors and appraisers (Johnson and Williams, 2021).

Additionally, the benefit of AI in real estate valuation is its capability to detect patterns and trends that evaluators might underestimate or overlook. As an example, AI could recognize subtle shifts in real estate market trends that could indicate potential price changes. This helps reduce the risks of undervaluation or overvaluation of real estate properties, which could significantly affect a company's financial statements (Brown, 2019), hence the work of auditors and accountants.

Even so, using AI in real estate valuation brings its own obstacle. While machine learning algorithms could deliver efficient and quick estimates, their clarity largely hinges on the quantity and quality of available and valid data. Moreover, AI models could be affected by systemic errors and biases, could be resulting in inaccurate valuations. Therefore, it is crucial for appraisers and auditors to identify the risks and limitations linked to these instruments and to supplement them with expertise and professional judgment in the real estate market. (Davis and Taylor, 2022).

Real estate valuation is a critical process with extensive applications across diverse fields, impacting both institutions and individuals. It serves a significant function in real estate transactions by establishing objective market prices and determining suitable rental rates for lease agreements (Büyükkaraciğan, 2021).

In this study, we center on advancing real estate valuation methods through the exploration of AVM. We examine how innovative techniques can improve the efficiency, accuracy, and overall efficacy of valuation processes. Our research highlights a comparative analysis of five Learning Systems techniques, including Decision Tree, Artificial Neural Networks, Linear Regression, Random Forest, and XGBoost. This process encompasses evaluating the performance of each technique using metrics, for example, the root mean square error which examines the variations between the actual values and predicted values (Hodson, 2022). By analyzing these methods, we aim to recommend new techniques that could revolutionize valuation practices and provide more scalable and reliable solutions for real estate, and for the related professions.

Table no. 3. Advanced techniques for real estate valuation

Category	Subcategory	Techniques
Cognitive Science Applications	*Learning Systems*	*Linear Regression*
		Decision Tree
		Random Forest
		Artificial Neural Networks (ANN)
		XGBoost

Source: Author's own composition

These methods are components of Learning Systems within Cognitive Science Applications.

The Cognitive Science Applications models and methods are designed to learn from the database and make accurate predictions (**Table no. 3**).

To be able to choose which of the methods proposed in the study is the most feasible and reliable for a certain region, the present study proposes a comparative analysis between the results obtained by each method. This step encompasses evaluating the performance of each technique using metrics, for example, the root mean square error (RMSE) which examines the variations between the actual values and predicted values (Hodson T. O. 2022).

3. Techniques in real estate valuation and their implications

3.1. Linear regression

Linear regression is a primary method for estimating quantitative results and, despite its historical longevity, remains one of the most efficient and extensively used techniques in statistics. While it might seem less advanced in contrast to other statistical approaches we will discuss further in this paper, linear regression is still a crucial instrument in data analysis. In addition, linear regression operates as a vital building block for more elaborate methods: contemporary statistical learning methods can be considered generalizations or extensions of this technique (James et al., 2023). This method has been successfully applied in other researches from specialized literature (Goundar S. et al. 2021, Sipos C. et al. 2008), thus succeeding in demonstrating its applicability in the real estate area.

Considering the complex nature of the real estate valuation process and the numerous attributes that influence property prices, relying exclusively on linear regression models for predictions is insufficient. To automate the valuation process and achieve accurate outcomes, it is mandatory to test multiple regression models. These techniques account for a wider range of influencing variables, providing a more reliable and thorough approach to assessing property values. The widespread popularity of multiple regression derives from its universal applicability to a variety of problems and data (Wang, 2003).

Linear regression is preferred for its robustness against violations of essential premises, its clear interpretation, and its broad availability through various statistical programs. These advantages make linear regression a go-to tool for analysts and researchers aiming to measure relationships between variables and create reliable predictions (Korkmaz, 2021).

In the example below, we will investigate how multiple regression can be used to estimate real estate value by considering several independent attributes such as the age of the building, the size of the house, the number of bedrooms, the number of rooms, the accessibility of the area, the city, the street, and the level of finish (Putra et al., 2023). The formula for our multiple regression model can be expressed as:

$$Y=\beta_0+\beta_1(\text{Number of Rooms})+\beta_2(\text{Size})+\beta_3(\text{Number of Bedrooms})+\beta_4(\text{City})+\beta_5(\text{Street})+\beta_6(\text{Accessibility})+\beta_7(\text{Age of the Building})+\beta_8(\text{Level of Finish})+\beta_9(\text{Lot Size})+\beta_{10}(\text{School Rating})+\beta_{11}(\text{Garage Size})+\beta_{12}(\text{Garden Size})+\beta_{13}(\text{Security Features})+\beta_{14}(\text{Energy Efficiency})+\ldots+\beta_n+\epsilon,$$

where:

Y – the dependent variable, representing the price of the house;

β_0 – the constant term; the expected value of price when all independent variables equal with zero;

$\beta_{1:n}$ – the column vector of the coefficients 1:n;

ϵ – the residual or error term, the variation in price not explained by the model.

It is noteworthy to mention that choosing the suitable attributes for the regression model is a fundamental procedure. This involves selecting only the significant attributes that have a significant effect on the price and ensuring a broad set of variables to accurately reflect the intricacies of the real estate market. In accordance with Heinze et al. (2018), several techniques can be applied in the attribute's selection process. These approaches include selecting attributes based on information criteria or significance, applying penalized likelihood, implementing background knowledge, utilizing the change-in-estimate criterion, or using a combination of these techniques. A thoughtfully chosen set of attributes helps increase the model's reliability and accuracy in predicting real estate values, validating that the outcomes reflect the various conditions in the market.

Nevertheless, linear regression has several significant weaknesses. It presumes a linear relationship between the independent and dependent variables, which may not always be accurate in practice, and it is very sensitive to outliers that can deform the result (Rousseeuw et al., 2005). In addition, it is the problem of multicollinearity, which can generate unreliable coefficient estimates. The approach also assumes independence of errors and homoscedasticity, assumptions that are often breached in practice. Moreover, linear regression can underfit or overfit data and struggles with complex datasets. It also presumes normally distributed residuals and deviations that can affect the confidence intervals and the hypothesis tests (James et al., 2013).

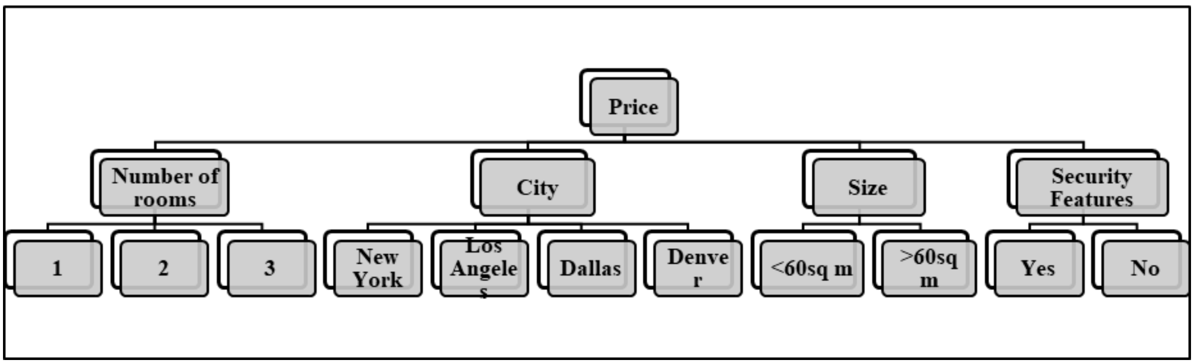
In the following part of the paper, we will explore more advanced estimation methods designed to overcome the weakness of the linear regression model.

3.2. Decision Tree

A decision tree is a guided learning model that structures a data domain into a hierarchical model, transposing it onto a set of outcomes. It iteratively divides the data domain into subdomains, ensuring that each split obtains a greater information gain than the prior node, leading to an increase in the power of prediction (Suthaharan, 2016).

For a better understanding of the process, we have illustrated in *Figure no.1* a simplified structure. A decision tree is a type of data organized in multiple nodes, each linked by branches. Nodes that have outbound edges are internal nodes, and the other ones are called leaves (Pekel, 2020). While this basic model assists in comprehending the fundamental structure of a decision tree, it is recommended, in practice, to utilize a higher number of variables to increase the accuracy of the prediction. The aim of *Figure no. 1* is to understand how various variables can impact the price, which is the target variable in this context.

Figure no. 1. Decision Tree in valuation process



Source: Author's own composition

At the root of the tree is price, the target variable, which we want to predict based on several influencing variables. The first level of branches splits the decision process into categories: the city of the property, the size of the property, the number of rooms, and the presence of security features. The first variable is the number of rooms with three possibilities: 1, 2, or 3 rooms. The second factor is the city where the property is located, branching into: New York, Los Angeles, Dallas, and Denver. The third variable is the size of the property, which divides into: properties smaller than 60 sq. m and properties bigger than 60 sq. m. After that, the 'security features' variable differentiates properties with and without security features Choudhury (Gupta et al., 2017).

This exemplified decision tree illustrates the hierarchical structure and the concept of decision trees as an education tool. By including more variables, one can build a more accurate and robust model for estimating target variables such as real estate values.

Decision trees are an efficient and accessible option for data analysis due to their simpleness. They are straightforward to visualize and understand and easy to interpret. In contrast to other methods that often require thorough data preparation, such as removing blank values, normalization, or creating dummy variables, decision trees necessitate a minimum level of preprocessing (Gupta et al., 2017). Furthermore, they generate accurate outcomes by employing measures such as Entropy, Gini index, and Information Gain to identify the optimal split at each node (Jadhav et al., 2016). These measures contribute to examining and selecting the best variables for dividing the data, ensuring that each split diminishes the impurity and maximizes the separation of classes in the dataset (Dash, 2022).

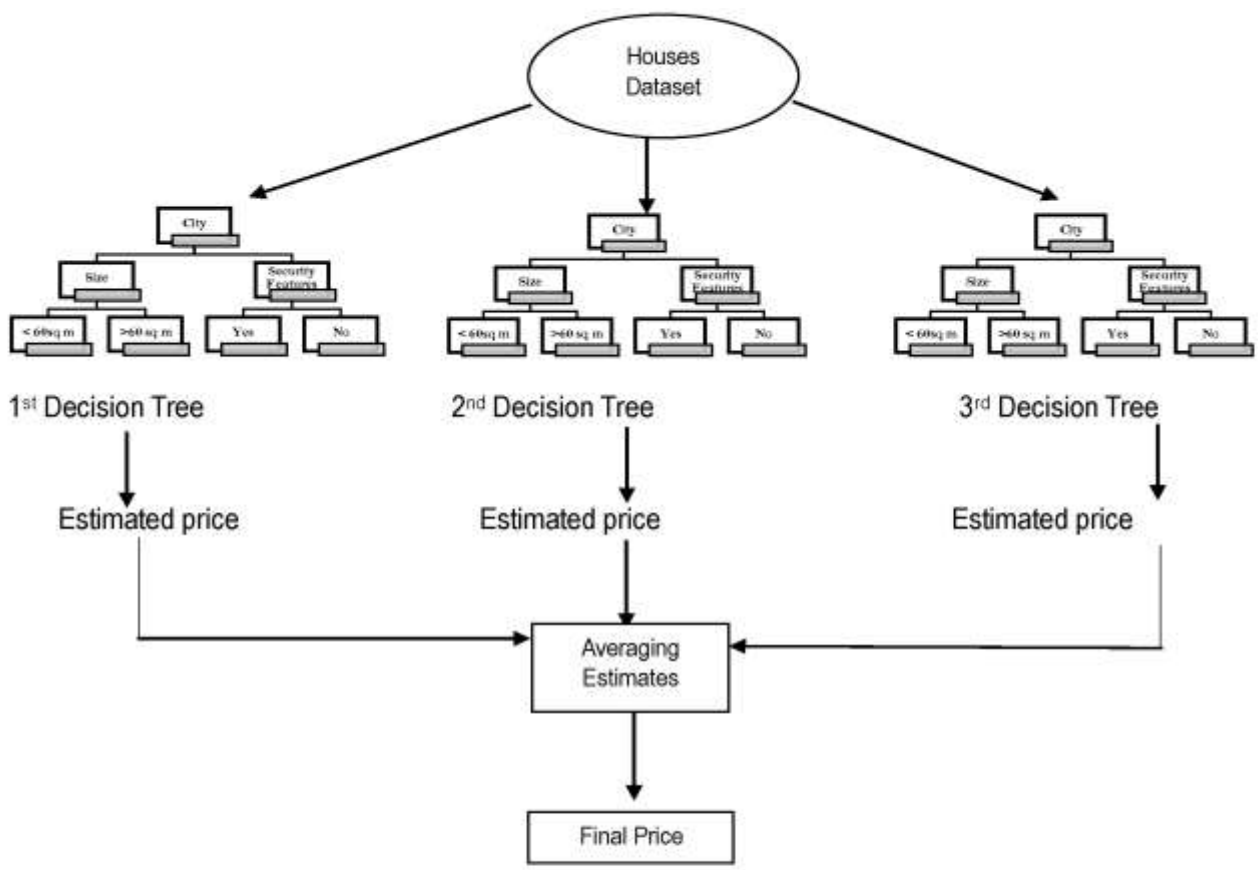
Despite their strengths, decision trees have several drawbacks. They are volatile, minimal data variations can notably alter the tree structure, and they are also susceptible to overfitting, detecting noise rather than underlying patterns, which decreases their generalizability and accuracy (Pehel, 2020). These limitations can influence the reliability of real estate valuations.

3.3. Random Forest

Developed by Breiman (2001), the random forest method has been demonstrated to be a highly effective tool for both regression and classification tasks. This algorithm works by generating multiple randomized decision trees and then merging their predictions through averaging (Biau et al., 2016). Every single decision tree node randomly picks a subset of factors from the entire dataset, and each tree utilizes a unique bootstrap sample of data, comparable to the bagging method (Oshiro et al., 2012). It excels, particularly in situations where the number of variables significantly surpasses the number of observations. Moreover, random forest is customizable to a broad range of extensive problems, easily adaptable for specific learning tasks, and supplies critical insights into variable significance.

To enhance the comprehension of the Random Forest algorithm, we will apply the same example previously utilized in the decision tree analysis. This approach will enable us to analyze and compare the methodologies underlying the advantages and specific features of Random Forest.

Figure no. 2. Random Forest in Valuation process



Source: Author's own composition

In the example from *Figure no. 2*, every decision tree in the Random Forest will autonomously evaluate the features of the property, such as the size, city, and security features, to estimate the target variable, the price. For example, one tree may focus on size and city, while another might prioritize security features and the size. This variation among the trees allows the Random Forest algorithm to capture a wide range of relationships and patterns within the data.

After all trees have made their individual property value predictions, these estimates are consolidated through averaging. By combining the estimations of multiple trees, Random Forest diminishes the risk of overfitting, which is a significant limitation of single decision trees. The averaging step also reduces the impact of biases or anomalies presented in individual trees, contributing to a more robust and accurate prediction.

Therefore, Random Forests is an enhanced version of a decision tree, applying multiple classifiers instead of one to improve the reliability and accuracy of predictions for upcoming instances (Shaik et al., 2019). Furthermore, it provides several advantages, including measuring the significance of each attribute in the training dataset, accurate predictions for a broad range of applications, and evaluating the pairwise distance between samples in the training data (Prajwala, 2015).

Nevertheless, the process of training Random Forest models can be highly resource-consuming, especially when dealing with extensive datasets and many trees. This requires significant processing power and memory, presenting a challenge for applications that necessitate real-time predictions (Hengl, 2018). Reproducing and validating Random Forest model outcomes can present challenges due to their complexity and randomness. Achieving reliable results requires maintaining the same model configurations and random seeds, which can be less transparent and burdensome than other techniques (Biau, 2012).

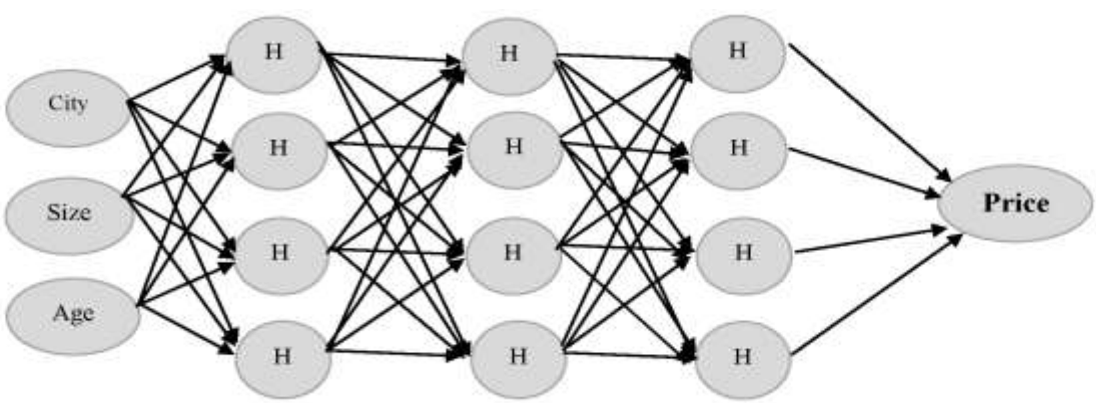
3.4. Artificial Neural Network

Artificial Neural Networks are a key topic in AI, inspired by the function and structure of the human brain. They model information processing and memory by generating elementary models that replicate the brain's neural networks. These models link diverse networks in different ways to process information similarly to the human brain (WU, 2018).

An ANN consists of interconnected neurons and each neuron can receive, process, and transmit signals. This network incorporates weighted synapses, which aggregate the input data according to these weights, and an activation mechanism that restricts the neuron's output amplitude, allowing the network to execute advanced computations by imitating the brain's neural process (Zhang, 2018).

Even though understanding an ANN can be complicated, we will keep it simple through a practical example. Explicitly, we will present how an ANN can predict property value based on three independent variables: size in sq. m, city, and age of the building. In practical predictions, including a larger number of attributes in a dataset is vital to ensure the accuracy of the prediction.

Figure no. 3. ANN in valuation process



Source: Author's own composition

In ***Figure no. 3***, the input layer of the artificial neural network includes three nodes representing the key variables that influence real estate values: city, size in sq. m, and age of the building. Every input node is equivalent to a specific variable of the property being evaluated. The first node captures the data about location. The size in sq. m node encodes the size of the property, and the age of the property node accounts for the building's age, which can influence its market value and condition. The ANN encompasses two hidden layers with several neurons marked with H. These hidden layers analyze the inputs through weighted connections and activation mechanism, capturing advanced non-linear relationships between the variables. The output layer, which is price is our example, merged the processed information to generate the predicted value of the property. This ANN structure models and predicts property value based on the specified attributes, utilizing the depth of the hidden layers to enhance the accuracy of the prediction.

One main advantage of the ANN algorithm is that it retains information across the entire network, instead of in one database. Therefore, losing information in one part of the network does not obstruct its overall functioning (Khalilov, 2021). The Artificial Neural Network feature’s superior fault tolerance and it's renowned for its high scalability and speed, especially when using parallel processing (Zou et al., 2009). It can manage binary inputs and outputs or symbolic data when it is correctly encoded, ensuring wide applicability across various domains (Wang S.C. et al. 2003). Moreover, they can learn from the environment, so they can be used for complex data or tasks where other types of solutions are impractical (Krenker, 2011).

Artificial Neural Networks have their own disadvantages, for example, the inclination to fall into local minima and the difficulty in adapting their architecture (Ding S. et al.2013). In addition, it can be challenging to fine-tune and optimize for specific assignments tasks (Abiodun et al., 2018). To enhance network generalization, it is necessary to utilize a network large enough to provide a suitable fit, as larger networks allow the creation of more elaborate functions (Dongare et al., 2012).

3.5. Extreme Gradient Boosting

Extreme Gradient Boosting is an advanced technique rooted in other boosting techniques like boosted classification trees and AdaBoost (Carmona et al., 2019). Extreme Gradient Boosting (XGBoost) can be utilized for both classification and regression problems and is preferred by data scientists for its out-of-core computation abilities and fast performance, making it suitable for capably managing large datasets (Osman et al., 2021). It utilizes a sparsity-aware mechanism that handles variables with missing entries or zero-values by automatically omitting these entries from the gain calculation for divided candidates, thus increasing the performance of the model (Bentéjac et al., 2021).

Applying the same example as in the previous techniques, we will illustrate the operation and the structure of XGBoost (***Figure no. 4***). The method starts by inputting the dataset, which includes the variables: city, size, and security features.

The first phase is to train the initial decision tree utilizing these attributes to initially estimate the property value. The variation between the estimated value and the actual value, identified as residuals, are then calculated, underlying the errors produced by the first tree.

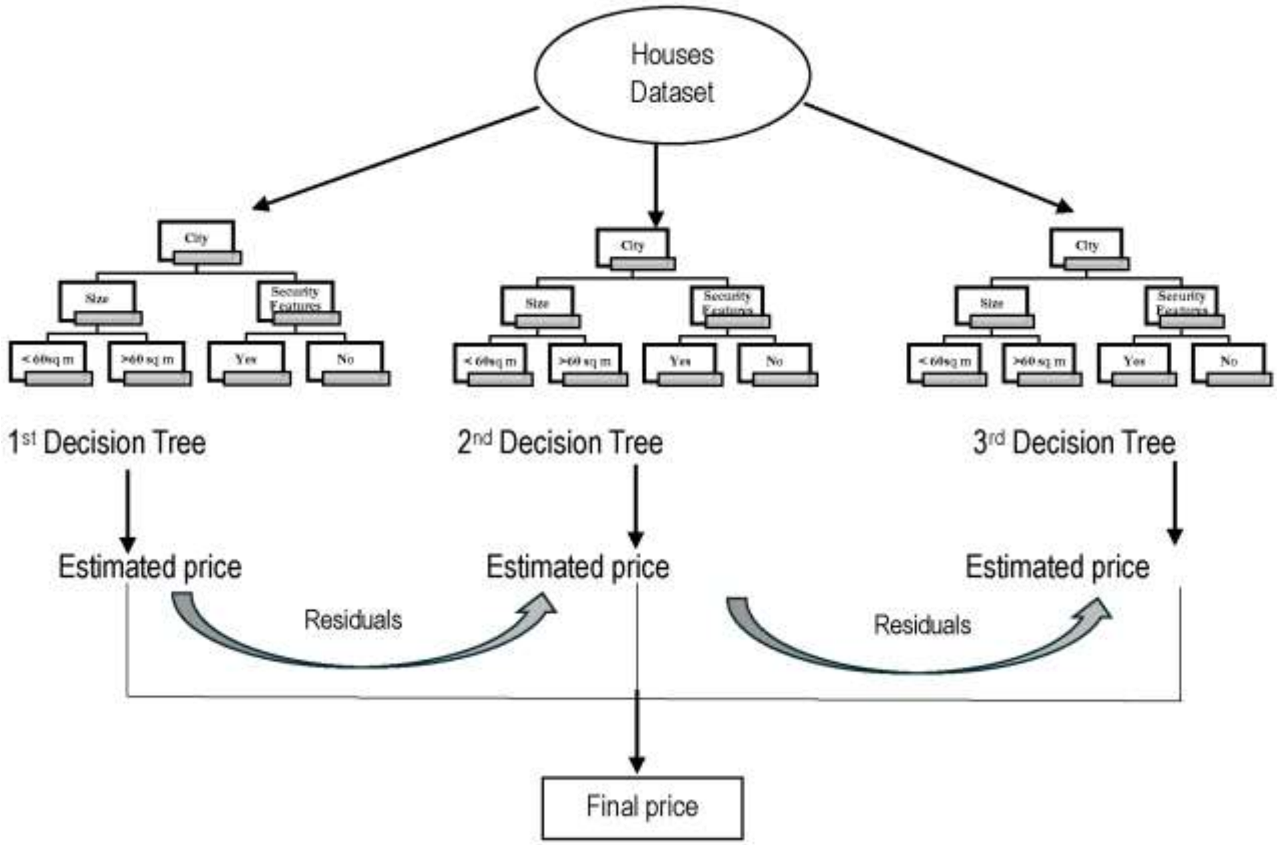
After that, the second decision tree is trained on these residuals with the aim of correcting the initial errors, thus increasing the accuracy of the model. This process of training successive trees and calculating residuals continues, and every new tree is focused on correcting the errors of the previous trees.

In the example, we included a third decision tree to enhance the value prediction by addressing the residuals from the second tree's estimations. The final property value prediction is derived by merging the outputs of all the decision trees. This repetitive process, named boosting, facilitates XGBoost to develop an accurate and performant model by constantly enhancing estimations through various stages of error correction, leading to an accurate value prediction.

The main features of XGBoost encompass its ability to manage sparse input for both linear and tree boosters, support for personalized objective and assessment functions, and consistent high performance across diverse datasets (Chen T. et al. 2015). Its success is due to the efficiency and scalability, as it runs at ten times the speed of other machine learning algorithms (Shilong, 2021). Moreover, XGBoost includes a regularization function within its aim to improve model generalization and avert overfitting (Zhou et al., 2021).

One limitation of the XGBoost algorithm is its disposition to overfit, which, if not properly managed, may result in but only average performance on the validation or test datasets, even if it has an exceptional performance on the training dataset (Drahokoupil, 2022).

Figure no. 4. Extreme Gradient Boosting in valuation process



Source: Author's own composition

Conclusion

In conclusion, the exploration of artificial intelligence techniques for real estate valuation exposes notable potential for changing traditional valuation. The real estate market, defined by its complexity and vulnerability to various factors, has the potential to benefit significantly from the implementation of AI-driven techniques. By applying advanced machine learning models, automated valuation techniques methods can accurately and quickly assess comprehensive datasets, offering consistent and reliable property valuation.

Artificial intelligence mechanisms can notably enhance the reliability of real estate valuations, reduce costs, and improve efficiency. Nevertheless, the efficiency of the AI techniques is dependent on the technical know-how of the professionals executing these models and on the quality of the data leveraged. Accordingly, it is crucial for real estate evaluators to ensure the high quality of the data and to acquire the necessary skills to implement AI methods efficiently.

In the paper, we examined the advantages and limitations of five AI-driven prediction methods that can be applied in real estate valuation. Each technique has its own weaknesses and strengths, and their performance can differ depending on the specific dataset. Considering the complexity and variation of real estate data, no unique technique secures the most effective results in every circumstance. Consequently, we highly recommend a thorough approach where all the presented AI methods are tested against the available dataset. Thus, professionals can practically identify which method generates the most accurate predictions. This process encompasses evaluating the performance of each technique using metrics, for example, the root mean square error which examines the variations between the actual values and predicted values. By rigorously evaluating and comparing all techniques, real estate experts can make data-driven decisions that improve the precision and reliability of property valuations.

In our opinion, it is essential that auditors, or the experts they rely on to verify the fair value and the appraisal report that forms the basis of the fair value estimate, are well-informed about automated valuations based on AI. Also, auditors must thoroughly understand how these AI-based models' function, including their methodologies, data inputs, and potential limitations to successfully implement them. This knowledge is critical to ensuring that the fair value reported in the financial statements is accurate, reliable, and transparent, thereby upholding the credibility of the financial information presented to investors and other stakeholders.

From our perspective, the contribution made by the use of automatic valuation methods is significant. In the first stage, within the implementation process, important time and financial resources are needed to gather data and compare all five proposed methods. But once the most suitable method has been selected, any property valuation can be done in a few seconds without generating additional costs. The only effort required by the appraiser is to enter the property's characteristics into the system, and in a few seconds, the program will automatically calculate the price. Therefore, once the evaluation method is implemented, substantial savings in both time and human resources will be recorded in practice, as well as an increase in accuracy.

Looking forward, sustained research and development in AI will be vital for further improving the performance and precision of real estate valuations. Upcoming research should investigate the challenges identified in the paper, such as data quality and model implementation, and examine new AI algorithms that could provide greater advantages. By welcoming these advancements, the real estate market can move towards more efficient, reliable, and scalable valuations, sustaining pricing decisions and investments.

Therefore, artificial intelligence is a revolutionary opportunity for the real estate market, yielding better consistency, accuracy, and efficiency. The knowledge acquired from this paper provides a solid foundation for ongoing exploration and implementation of AI in real estate valuation, facilitating progress toward a future where automated valuation methods become fundamental to the standard procedures.

References

1. Abiodun O. I., Jantan A., Omolara A. E., Dada K. V., Mohamed N. A., & Arshad H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11)
2. Aji N. A., & Dhini A. (2019, July). Credit scoring through data mining approach: A case study of mortgage loan in Indonesia. *2019, 16th International Conference on Service Systems and Service Management* (ICSSSM), pp. 1-5
3. Bentéjac C., Csörgő A., & Martínez-Muñoz G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, 1937-1967
4. Bhatnagar R., & Gohain G. B. (2020). Crop yield estimation using decision trees and Random Forest machine learning algorithms on data from terra (EOS AM-1) & Aqua (EOS PM-1) satellite data. *Machine Learning and Data Mining in Aerospace Technology*, 107-124
5. Biau G. (2012). Analysis of a Random Forests model. *The Journal of Machine Learning Research*, 13(1), 1063-1095
6. Biau G., & Scornet E. (2016). A Random Forest guided tour. *Test*, 25, 197-227
7. Boztosun D., Aksoylu S., & Ulucak, Z. Ş. (2016). The role of human capital in economic growth. *Economics World*, 4(3), 101-110
8. Breiman L. (2001). Random Forests. *Machine learning*, 45, 5-32
9. Briganti G., & Le Monie O. (2020). Artificial Intelligence in Medicine: Today and Tomorrow. *Frontiers in Medicine*, (7)27
10. Brown, R. (2019). Evaluating Market Conditions in Real Estate Valuation. *Real Estate Appraisal Journal*, 56(2), 75-88.
11. Brown, R. (2019). The Role of IA in Modern Real Estate Markets. *Real Estate Economics*, 48(2), 123-140.

12. Brown, R. (2020). Understanding Automated Valuation Models in Real Estate. *Real Estate Appraisal Journal*, 58(1), 90-104.
13. Büyükkaracığan D. N. (2021). Modern methods approach in real estate valuation. *Ankara: Iksad Publishing House*, 130
14. Carmona P., Climent F., & Momparler A. (2019). Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 61, 304-323
15. Chen T., He T., Benesty M., Khotilovich V., Tang Y., Cho H., ... & Zhou T. (2015). XGboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 1-4
16. Choudhury M. (2015). Subjective probability and financial valuation: contrasting paradigms. *Journal of Financial Reporting and Accounting*, 13(1), 20-38
17. Dash S. (2022). Decision Trees Explained – Entropy, Information Gain, Gini Index, CCP Pruning. *Towards Data Science*, 46
18. Davis, M. & Taylor, S. (2022). The Intersection of Valuation and Audit in Financial Reporting. *Journal of Accounting and Economics*, 60(3), 215-230.
19. Davis, M., & Williams, K. (2022). Transparency in Fair Value Reporting: The Role of IA-Based Valuations. *International Journal of Financial Reporting*, 29(3), 150-167.
20. Ding S., Li H., Su C., Yu J., & Jin F. (2013). Evolutionary artificial neural networks: a review. *Artificial Intelligence Review*, 39, 251-260
21. Dongare A. D., Kharde R. R., & Kachare A. D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology* (IJEIT), 2(1), 189-194
22. Drahokoupil J. (2022). Application of the XGBoost algorithm and Bayesian optimization for the Bitcoin price prediction during the COVID-19 period. *FFA Working Papers*
23. Ferrari R. (2015). Writing narrative style literature reviews. *Medical writing*, 24(4), 230-235
24. Ge Y., & Wu H. (2020). Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Computing and Applications*, 32, 16843-16855
25. Ghosh A., & Maiti R. (2021). Soil erosion susceptibility assessment using logistic regression, decision tree and Random Forest: study on the Mayurakshi river basin of Eastern India. *Environmental Earth Sciences*, 80(8), 328
26. González C., Mira J. M., & Ojeda J. A. (2016). Applying multi-output Random Forest models to electricity price forecast, *Statistical Laboratory, Technical University of Madrid*
27. Goundar S., Maharaj K., Kumar A., & Bhardwaj A. (2021). Property valuation using linear regression and Random Forest algorithm. *International Journal of System Dynamics Applications* (IJSDA), 10(4), 1-16
28. Gupta B., Rawat A., Jain A., Arora A., & Dhami N. (2017). Analysis of various decision tree algorithms for classification in data mining. *International Journal of Computer Applications*, 163(8), 15-19
29. Heinze G., Wallisch C., & Dunkler D. (2018). Variable selection–a review and recommendations for the practicing statistician. *Biometrical journal*, 60(3), 431-449
30. Hengl T., Nussbaum M., Wright M. N., Heuvelink G. B., & Gräler B. (2018). Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518
31. Hodson T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 2022, 1-10
32. Huang J., Saleh S., & Liu Y. (2021). A Review on Artificial Intelligence in Education. *Academic Journal of Interdisciplinary Studies*, 10(3). 207-217
33. Jadhav S. D., & Channe H. P. (2016). Comparative study of K-NN, naive Bayes and decision tree classification techniques. *International Journal of Science and Research* (IJSR), 5(1), 1842-1845
34. James G., Witten D., Hastie T., & Tibshirani R. (2013). An introduction to statistical learning (Vol. 112, p. 18). *New York: Springer*.
35. James G., Witten D., Hastie T., Tibshirani R., & Taylor J. (2023). Linear regression. In An introduction to statistical learning: With applications in python. *Springer International Publishing*, 69-134
36. Johnson, L. & Williams, K. (2021). Financial Reporting and the Role of Fair Value in Real Estate. *International Journal of Financial Auditing*, 33(4), 589-603.

37. Johnson, L. & Williams, K. (2021). Machine Learning in Real Estate: Implications for Auditors. *International Journal of Financial Auditing*, 32(4), 567-580.
38. Jha, G. K., & Sinha, K. (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. *Agricultural Economics Research Review*, 26(2), 229-240
39. Khaidem L., Saha S., & Dey S. R. (2016). Predicting the direction of stock market prices using Random Forest. *arXiv preprint arXiv:1605.00003*.
40. Khalilov D. A., Jumaboyeva N. A. K., & Kurbonova T. M. K. (2021). Advantages and Applications of Neural Networks. *Academic research in educational sciences*, 2(2), 1153-1159
41. Khan Z. (2022). Used car price evaluation using three different variants of linear regression. *International Journal of Computational and Innovative Sciences*, 1(1), 12-20
42. Khan Z. H., Alin, T. S., & Hussain, M. A. (2011). Price prediction of share market using artificial neural network (ANN). *International Journal of Computer Applications*, 22(2), 42-47
43. Khanzode K. C. A., & Sarode R. D. (2020). Advantages and disadvantages of artificial intelligence and machine learning: A literature review. *International Journal of Library & Information Science* (IJLIS), 9(1), 3
44. Korkmaz M. (2021). A study over the general formula of regression sum of squares in multiple linear regression. *Numerical Methods for Partial Differential Equations*, 37(1), 406-421
45. Krenker A., Bešter J., & Kos A. (2011). Introduction to the artificial neural networks. Artificial Neural Networks: Methodological Advances and Biomedical Applications. *InTech*, 1-18
46. Langsetmo L., Schousboe J. T., Taylor B. C., Cauley J. A., Fink H. A., Cawthon, P. M., ... & Osteoporotic Fractures in Men (MrOS) Research Group. (2023). Advantages and disadvantages of Random Forest models for prediction of hip fracture risk versus mortality risk in the oldest old. *Journal of Bone and Mineral Research Plus*, 7(8)
47. Lasota T., Łuczak, T. Niemczyk M., Olszewski M., & Trawiński B. (2013). Investigation of property valuation models based on decision tree ensembles built over noised data. In Computational Collective Intelligence. *Technologies and Applications: 5th International Conference*, ICCCI 2013, Craiova, Romania, September 11-13, 2013, Proceedings 5 (pp. 417-426). Springer Berlin Heidelberg
48. Ma B., Meng, F., Yan G., Yan H., Chai B., & Song F. (2020). Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. *Computers in biology and medicine*, 121
49. Manoj, J., & Suresh, K. K. (2019). Forecast model for price of gold: Multiple linear regression with principal component analysis. *Thailand Statistician*, 17(1), 125-131
50. Mombeini H., & Yazdani-Chamzini A. (2015). Modeling gold price via artificial neural network. *Journal of Economics, business and Management*, 3(7), 699-703
51. Mohammad S. J., Hamad A. K., Borgi H., Thu P. A., Sial M. S., & Alhadidi A. A. (2020). How artificial intelligence changes the future of accounting industry. *International Journal of Economics and Business Administration*, 8(3), 478-488
52. NandigalaVenkatAnurag, Y., & Sharanya, S. (2019). Air Quality Index Prediction with Meteorological Data Using Feature Based Weighted Xgboost. *International Journal of Recent Technology and Engineering* (IJRTE), 8(1), 1355-1358.
53. Nikolaev D., & Petrova M. (2021, October). Application of simple convolutional neural networks in equity price estimation. In *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology* (PIC S&T), pp. 147-150
54. Nwulu, N. I. (2017, September). A decision trees approach to oil price prediction. In *2017 International Artificial Intelligence and Data Processing Symposium* (IDAP), pp. 1-5
55. Oba, K. M. (2019). A multiple linear regression model to predict the price of cement in Nigeria. *International journal of economics and management engineering*, 13(12), 1480-1485
56. Oshiro T. M., Perez P. S., & Baranauskas J. A. (2012). How many trees in a Random Forest?. *In Machine Learning and Data Mining in Pattern Recognition: 8th International Conference*, MLDM 2012, Berlin, Germany, July 13-20, 2012.

Proceedings 8 (pp. 154-168). Springer Berlin Heidelberg

57. Osman A. I. A., Ahmed A. N., Chow M. F., Huang Y. F., & El-Shafie A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, 12(2), 1545-1556
58. Padmanaban, K. A., & Parthiban, G. (2016). Applying machine learning techniques for predicting the risk of chronic kidney disease. *Indian Journal of Science and Technology*, 9(29), 1-6
59. Pannu A. (2015). Artificial intelligence and its application in different areas. *Artificial Intelligence*, 4(10), 79-84
60. Patel, A. (2023). Integrating IA with Traditional Real Estate Valuation Techniques. *Global Real Estate Journal*, 29(5), 345-362.
61. Pekel E. (2020). Estimation of soil moisture using decision tree regression. *Theoretical and Applied Climatology*, 139(3), 1111-1119
62. Prajwala T. R. (2015). A comparative study on decision tree and Random Forest using R tool. *International journal of advanced research in computer and communication engineering*, 4(1), 196-199
63. Putra P. H., Azanuddin A., Purba B., & Dalimunthe Y. A. (2023). Random Forest and decision tree algorithms for car price prediction. *Jurnal Matematika Dan Ilmu Pengetahuan Alam LLDikti Wilayah 1* (JUMPA), 3(2), 81-89
64. Ramani K., Jahnavi M., Reddy P. J., VenkataChakravarthi, P., Meghanath P., & Imran S. K. (2023, March). Prediction of bitcoin price through LSTM, ARIMA, XGBoost, prophet and sentiment analysis on dynamic streaming data. In *2023 9th International Conference on Advanced Computing and Communication Systems* (ICACCS) (Vol. 1, pp. 1514-1518)
65. Rousseeuw P. J., & Leroy A. M. (2005). Robust regression and outlier detection. *John wiley & Sons*
66. Roy, S. S., Mittal, D., Basu, A., & Abraham, A. (2015). Stock market forecasting using LASSO linear regression model. In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014* (pp. 371-381)
67. Saini, D., Saxena, A., & Bansal, R. C. (2016, December). Electricity price forecasting by linear regression and SVM. In *2016 International Conference on Recent Advances and Innovations in Engineering* (ICRAIE), pp. 1-7
68. Shaik A. B., & Srinivasan S. (2019). A brief survey on Random Forest ensembles in classification model. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018*, Volume 2, pp. 253-260
69. Shanbehzadeh, M., Nopour, R., & Kazemi-Arpanahi, H. (2022). Design of an artificial neural network to predict mortality among COVID-19 patients. *Informatics in medicine unlocked*, 31
70. Shilong Z. (2021). Machine learning model for sales forecasting by using XGBoost. In *2021 IEEE international conference on consumer electronics and computer engineering*, pp. 480-483
71. Sisodia D., & Sisodia D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science*, 132, 1578-1585
72. Smith, J. (2020). IA and Real Estate: The Future of Property Valuation. *Journal of Real Estate Research*, 45(3), 221-234.
73. Smith, J. (2020). Theoretical Foundations of Property Valuation. *Journal of Real Estate Economics*, 42(1), 112-129.
74. Svetlana N., Anna N., Svetlana M., Tatiana G., & Olga M. (2022). Artificial intelligence as a driver of business process transformation. *Procedia Computer Science*, *213*, 276-284.
75. Suthaharan S. (2016). Decision tree learning. Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, *Integrated Series in Information Systems*, 36, 237-269
76. Şipoş C., Crivii E. A., & FRICS M. (2008). A linear regression model for real estate appraisal. In *WAVO Valuation Congress Valuation in Diversified and Emerging Economies*, pp. 17-18
77. Taylor, S. (2023). Ensuring Accuracy in IA-Driven Valuations for Financial Reporting. *Journal of Accounting and Economics*, 62(1), 35-50.
78. Vaiz J. S., & Ramaswami M. (2016). A study on technical indicators in stock price movement prediction

using decision tree algorithms. *American Journal of Engineering Research* (AJER), 5(12), 207-12

79. Ugurlu U., Oksuz I., & Tas O. (2018). Electricity price forecasting using recurrent neural networks. *Energies*, 11(5), 1255
80. Vuong P. H., Dat T. T., Mai T. K., & Uyen P. H. (2022). Stock-price forecasting based on XGBoost and LSTM. *Computer Systems Science & Engineering*, 40(1)
81. Wang S. C. (2003). Artificial neural network. *Inter-disciplinary computing in java programming*, 81-100
82. Wu Y. C., & Feng J. W. (2018). Development and application of artificial neural network. *Wireless Personal Communications*, 102, 1645-1656
83. Ya, K., Li W., Ji Z., Qi M., & Du Y. (2019). A hybrid LSTM neural network for energy consumption forecasting of individual households. *Ieee Access*, 7, 157633-157642
84. Zhang Z., (2018). Artificial neural network. *Multivariate time series analysis in climate and environmental research*, 1-35
85. Zou J., Han Y., & So S. S. (2009). Overview of artificial neural networks. *Artificial neural networks: methods and applications*, 14-22
86. Zhou J., Qiu Y., Zhu S., & Jahed D. (2021). Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization. *Underground Space*, 6(5), 506–515
87. Zhou, L., Pan S., Wang J., & Vasilakos A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237, 350-361
88. Zhou Y., Li, T., Shi J., & Qian Z. (2019). A CEEMDAN and XGBOOST-based approach to forecast crude oil prices. *Complexity*, 2019(1)
89. Zhou T., Shi P., Jia G., & Luo Y. (2013). Nonsteady state carbon sequestration in forest ecosystems of China estimated by data assimilation. *Journal of Geophysical Research: Biogeosciences*, 118(4), 1369-1384