



HAL
open science

Machine Learning Applications in Real Estate: Critical Review of Recent Development

Jamal Al-Qawasmi

► **To cite this version:**

Jamal Al-Qawasmi. Machine Learning Applications in Real Estate: Critical Review of Recent Development. 18th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Jun 2022, Hersonissos, Greece. pp.231-249, 10.1007/978-3-031-08337-2_20 . hal-04668652

HAL Id: hal-04668652

<https://inria.hal.science/hal-04668652v1>

Submitted on 7 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Machine Learning Applications in Real Estate: Critical Review of Recent Development

Jamal Al-Qawasmi [0000-0002-4708-8738]

Architecture Department, King Fahd University for Petroleum & Minerals, Dhahran 31261,
Saudi Arabia

jama1q@kfupm.edu.sa

Abstract. Machine learning (ML) and deep learning (DL) methods have recently become a hot topic in the real estate discipline. They have contributed to the advancement of various domains in real estate sector. This paper provides a critical review of recent trends in applying machine learning and deep learning (ML/DL) techniques in various domains of real estate and investigate their potential for the real estate sector. Recent advances in model development, testing and areas of application in real estate in the past 4 years (2017-2020) are presented. Findings reveal that 20 different ML & DL algorithms were utilized to examine various aspects of real estate development and valuation, and that the most commonly used algorithms are neural networks, regression models, random forest, booting, support vector machine and cubist /pruned model tree.

Keywords: machine learning, deep learning, real estate, properties.

1 Introduction

Real estate is the arena of human activities. During the past two decades or so, machine learning (ML) and deep learning (DL) algorithms and methods have moved into real estate research, thus there is enormous opportunities for future research in order to enhance several areas in the field. ML/DL algorithms and models have been applied in a wide range of activities and processes in real estate development and valuation such as valuation of land, mass appraisal for taxation purposes, renting value, and market value of real estate properties. However, the application of ML and DL in this field still new and more growth is expected in the future [1].

Valuation is one of the major application areas of ML/DL in real estate. Real estate valuation is the process of developing an opinion of value for real estate property. Real estate valuation and estimations are needed for purposes such as selling, buying, renting, real estate development, secure bank loans, taxation, amongst others. In all these cases, stakeholders are interested in getting the most accurate estimation of the value of the real estate property. Traditionally, the valuation of real estate properties is based on the expertise of the valuers/appraisers. This process suffers from inaccuracy in estimating the value of proprieties due to the differences in the perception of valuers.

As such, the conventional valuation process is often supported by computer systems called Automated Valuation Models (AVM) and Computer Assisted Mass Appraisal (CAMA) which employ various statistical methods and models to estimate the value of a single or a mass of properties [2], [3]. The ML/DL methods have entered the field as new addition to these conventional practices.

This paper aims to identify the main ML and DL techniques used in the real estate field. The paper provides an extensive survey of recent developments and achievements in applying ML/DL algorithms and techniques to advance real estate field. Extensive literature review showed that there is no similar systematic survey on the topic in the past decade or so [4].

This survey paper is organized as follows: section 2 summarizes the paper methodology. Section 3 examines and analyzes the ML and DL algorithms and techniques applied in real estate literature and outlines the relevant studies, section 4 provides a discussion of results, and section 5 provides conclusion of findings.

2 Materials and Methodology

To identify relevant literature in the area of applying ML/DL algorithms and techniques in real estate development and valuation, the author explored Google Scholar, Web of Science (WoS), and Scopus databases with the following search keywords: “real estate”, “property” or “mass appraisal” along with “machine learning”, “deep learning” in addition to names of a wide range of ML and DL algorithms. The search in these databases showed that there is high increase in using ML and DL algorithms in real estate development and valuation during the past decade or so. However, the application of ML and DL in this field still new and more growth is expected in the future. Our search was focused on recent research studies that have been published in the past four years (i.e. between 2017 and 2020). After removing duplicates, the search resulted in a total number of 36 papers. Based on reading the abstract of each paper, a total of 16 papers have been excluded as they are unrelated to the topic. We also excluded most of conference papers as our research is focused on journal papers. Only 20 papers were remained for analysis, for of them are conference papers. The final list of papers selected for analysis is presented in Table 1.

Table 1. Descriptive analysis of the 22 examined research papers.

Study	Pub. year	Study area	ML/DI methods	Objective of the study
Abidoye & Chan [5]	2017	Lagos, Nigeria	Artificial Neural Network	Examine accuracy and feasibility of using ANN technique to predict the prices of properties
Guo et al. [6]	2019	Beijing, & other 5 cities, China	Artificial Neural Network	Optimize ANN to work for cities with insufficient real estate data without compromising accuracy.
Alexandridis et al. [7]	2019	Several cities, Greece	Artificial Neural Network	Assess performance & accuracy of ANN models with 2

				traditional real estate valuation methods (i.e MRA & SMV)
Abidoye & Chan [8]	2018	Lagos metropolis, Nigeria	Artificial Neural Network	Evaluate the predictive accuracy of ANN model with the hedonic pricing model (HPM) in residential property valuation.
Yu et al. [9]	2018	Beijing, China	Convolution Neural Network (CNN), Recurrent Neural Network (RNN), Logical regression model	Comparative analysis of prediction power of 3 ML models of the market value of second-hand housing
Poursaeed et al. [10]	2018	(House photos from) USA	Convolutional neural network	Predict the impact of (quality of) appearance on the value of residential properties using CNN.
Bin et al. [11]	2017	NA	Recurrent Neural Network (RNN)	Apply RNN in a model to predict property value & compare its accuracy/performance with other non-ML appraisal models
Petkov [12]	2020	Lisbon, Portugal	Self-organizing Map (SOM), & Support Vector Machine (SVM)	Compare performance of SOM with SVM, & hedonic price model (HPM) in predicting rental prices.
Ceh et al. [13]	2018	Ljubljana, Slovenia	Random Forest	Empirically evaluate the predictive power of the random forest technique compared with the hedonic pricing model (HPM) in predicting of prices apartments.
Aydinoglu et al. [14]	2020	Istanbul, Turkey	Random Forest	Propose a land valuation model for residential properties as an extension of the national geographic data infrastructure (GDI) and assess its accuracy in estimating market value of land.
Dimopoulos et al. [15]	2018	Nicosia, Cyprus	Random Forest, Linear (Multivariate) Regression	Compare the performance of random forest method with linear multivariate regression in predicting apartments values.
Alfaro-Navarro et al. [16]	2020	48 provinces in Spain	Boosting, Random forest, and Bagging	Automate the selection of best ML/DL model for each city of a country, & compare the prediction power of 3 ML/DL models of the market value of properties, and identify variables that impact value.
Kok et al. [17]	2017	California, Florida, and Texas, USA	Decision tree	Evaluate the accuracy of decision tree models compared to the hedonic pricing model

				(HPM) in predicting value and net operation income (NOI) of multi-family properties.
Dimopoulos and Bakas [18]	2019	Nicosia, Cyprus	Gradient Boosting, Random Forests, Linear Regression, and Non-Linear Regression	empirically investigate how the four ML/DL models work to predict apartment prices, and how it takes various variables (property characteristics) into consideration.
Vargas-Calderón & Camargo [19]	2019	Bogota, Colombia	XGBoost (gradient boosting)	Investigate the ability of XGBoost, a gradient boosting algorithm, to predict property prices based on data published in real estate market websites
Gružasuskas et al. [20]	2020	Tauragė, Lithuania	Boosted Decision Tree	Develop and test a model based on the boosted decision tree algorithm as a supplement to traditional “comparative method” when similar historical data is unavailable or limited.
Bogdan et al. [21]	2017	Poland	Pruned Model Tree (M5P), Multilayer Perceptron, Linear Regression	Evaluate the accuracy of predicting sales prices of residential building using 3 ML/DL models (M5P, MLP & linear regression) compared with 2 expert algorithms used in real state valuation.
Trawiński et al. [22]	2017	Wrocław, Poland	Pruned Model Tree (M5P), Multilayer Perceptron	Evaluate the accuracy of predicting sales prices of apartments by Pruned Model Tree (M5P), and Multilayer Perceptron (MLP) algorithms compared with 3 expert algorithms used in real state valuation.
Clark and Lomax [23]	2018	England (whole country)	Cubist, Gradient Boost, Multivariate adaptive regression splines (MARS), Support Vector Machine (SVM)	Assess performance & accuracy of 4 ML/DL algorithms (Cubist, gradient boosting, MARS, & SVM) and one hedonic price model (GLM).
Derdouri & Murayama [24]	2020	Fukushima, Japan	Generalized linear model (GLM), Generalized additive model using splines (GAMS), Multivariate adaptive regression spline (MARS), k-nearest neighbors (kNN), SVMRadial, Cubist, Stochastic gradient boosting (GBM), Random forest.	compare and evaluate the performance of 9 ML/DL algorithms and one non-ML algorithm in predicting land prices.

3 ML/DL Algorithms and their Applications in Real Estate

The analyzed 20 research papers have proposed and tested several ML/DL algorithms for application in the real estate field. A total of 20 different ML/DL algorithms have been examined in the 20 papers. The number of ML/DL algorithms examined in each paper ranged between one to nine. The following is a detailed analysis of each paper, the ML/DL algorithms examined, and the major findings.

3.1 Neural Networks (NN)

Neural networks have a wide range of applications in real estate. They are the most used ML/DL methods in real estate. As shown in Table 1, ten research papers (out of 20 papers) have addressed the use of 11 neural network models in the context of real estate development and valuation. Five types of neural network algorithms were examined: artificial neural network (ANN), convolutional neural network (CNN), recurrent neural network (RNN), multilayer perceptron (MLP) and self-organizing map (SOM).

Artificial neural network (ANN). Abidoye & Chan [5], for instance, investigated the accuracy and feasibility of using ANN technique to predict the prices of properties in Lagos, Nigeria. The study examined the impact of 11 property attributes on the value of property. Results showed that ANN provided suitable and reliable predictions of values of properties. The model also able to assess the impact of each attribute on the property value.

Guo et al. [6] developed a framework to predict prices of properties across several cities in china using semi-supervised ANN. This is done by separating homogeneous attributes from heterogeneous attributes of properties and dealing with each separately.

Alexandridis et al. [7], on the other hand, proposed an ANN model to forecast the value of properties in Greece and compared its performance and accuracy with traditional real estate valuation methods; the multiple regression analysis (MRA) and similarity measure valuation (SMV) models. After examining the strengths and the performance of each method, authors applied a combined forecasting rule to improve forecasting accuracy. They also identified the property characteristics that lead to large forecasting errors. Results showed that the proposed ANN model provides an accurate tool for property valuation in non-homogeneous, newly developed markets, and constantly outperform traditional valuation methods (i.e. MRA and SMV).

Abidoye & Chan [8] evaluated the predictive accuracy of hedonic pricing model (HPM) in comparison with the artificial neural network (ANN) technique in property valuation. Residential property transaction data were collected from registered real estate firms operating in the Lagos metropolis, Nigeria, and were fitted into the ANN and HPM models. The results showed that the ANN technique outperformed the HPM approach in terms of accuracy in predicting property values with higher r^2 and lower accuracy error (MAPE, RMSE, MAPE).

Convolutional neural network (CNN) and Recurrent Neural Network (RNN). Yu et al. [9] built three ML/DL models, namely Convolution Neural Network (CNN), Recurrent Neural Network (RNN) and logistic regression model, to predict housing prices in a city according to range of characteristics such as building age, location,

surrounding facilities and so on. They found that the accuracy of the prediction of a model is directly affected by the characteristic and attributes used. For example, they found that CNN is more accurate in predicting when using the dataset based on characteristic factors obtained by deep crawling, while the RNN is more accurate in predicting when using the time series data.

Poursaeed et al. [10] have used convolutional neural network (CNN) in a novel application to evaluate the impact of visual characteristics of residential properties (i.e. their interior and exterior appearance) on their market value. They develop a method that used deep convolutional neural networks on a large dataset of photos of home interiors and exteriors for estimating the luxury level of real estate photos and estimate their value. They found that the approach is more accurate and provide better value estimation compared to existing methods that do not consider visual aspects of residential buildings.

Bin et al. [11] proposed a boosting tree model facilitated with a recurrent neural network (RNN) to forecast the average price of real estate in a city. The model uses properties attributes such as qualities of houses, location, and trend of market price to estimate property value. However, the model doesn't take spatial and temporal aspects of properties into account. The results indicate that the proposed RNN model outperforms the existing non-ML models used in the real estate appraisal industry.

Self-organizing map (SOM). For example, Petkov [12] used self-organizing map (SOM), a type of ANN, and support vector machine (SVM) to predict mid-term rental prices based on various spatial variables.

Multilayer perceptron (MLP). MLP is a special type of feedforward artificial neural network. Each of Bogdan et al. [21] and Trawiński et al. [22] conducted a comparative study in which they examined the performance of MLP method compared to other ML algorithms. These two papers are discussed in section 3.4 below.

Table 2. Applications of neural networks in real estate and major findings. *

Study	Highlighted finding	Domain of application
Abidoye & Chan [5]	Price predictions of ANN technique are feasible, highly accurate and reliable.	Property valuation, mass appraisal, predict property value.
Guo et al. [6]	The developed semi-supervised ANN model achieved similar or even superior performance compared to a fully supervised ANN in property evaluation.	Property valuation, mass appraisal, predict property value.
Alexandridis et al. [7]	The proposed ANN method is an accurate prediction tool and constantly outperform traditional valuation methods (i.e. MRA and SMV).	Property valuation in non-homogeneous, newly developed markets.
Abidoye & Chan [8]	The proposed ANN method outperformed the HPM approach in prediction of market value for residential properties.	Property valuation of residential properties, mass appraisal.
Yu et al. [9]	Accuracy of ML/DL methods (CNN, RNN & LRM) depends on type of attributes/variables used. CNN is more accurate when using	Market value of second-hand housing.

	housing characteristic data, while RNN is more accurate when using time series data.	
Poursaeed et al. [10]	The approach (CNN & novel algorithm) is more accurate and provide better value estimation compared to existing methods that do not consider visual aspects of residential buildings.	Valuation of various types of residential buildings based on quality/luxury level of property.
Bin et al. [11]	The proposed RNN model outperformed the existing non-ML models used in the real estate appraisal industry.	Property valuation, property appraisal.
Petkov [12]	ML methods (SVM and SOM) outperformed the non-ML methods (i.e. hedonic price model). Accuracy of ML methods depends on type of attributes/variables used.	Mid-term rental prices.

* See also papers [21] & [22] that address these algorithms

3.2 Random Forest, Bagging and Decision Tree

Six papers have investigated the application of random forest algorithm, while one paper has examined each of bagging and decision tree algorithms, as shown in Table 3. For instance, Ceh et al. [13] examined the predictive performance of the random forest algorithm in comparison to commonly used hedonic pricing models (HPM) in predicting prices of apartments in Slovenia. Results revealed that prediction results of the random forest model are significantly better than the results obtained through HPM model according all performance measures (R2 values, sales ratios, mean average percentage error (MAPE), coefficient of dispersion (COD)). Ten of the most important predicting variables identified by random forest models are also supported by Principal Component Analysis (PCA).

Aydinoglu et al. [14] proposed an interoperable land valuation model using random forest method and evaluated its accuracy in estimating market value of residential urban land. The model is developed as an extension of the national geographic data infrastructure (GDI) of Turkey. Experimental results showed that the random forest algorithm accurately predicted the value of residential land in the light of all related performance measures (i.e. median ratio, COD and PRD) and standards International Association of Assessing Officers (IAAO).

Dimopoulos et al. [15] conducted a comparative study to assess the performance and accuracy of random forest and linear multivariate regression methods in predicting market values of apartments (multifamily residences). The result revealed that the random forest method outperformed the linear multivariate regression in predicting market value of apartments, and that the difference in the accuracy measure (the root mean squared error -RMSE) between the two models ranged between 7.87% and 11.17%. The set of most important variables affected the value of apartments were identified.

Alfaro-Navarro et al. [16] developed an application that automates the selection of best ML model for each city/district in Spain. Real estate property prices in 433 municipalities are estimated from a sample of 790,631 dwellings, using three ensemble methods (bagging, random forest and boosting) based on decision trees. Results showed that the ability of the three ensemble methods (bagging, random forest and

boosting) in estimating the price of dwellings is clearly outperformed individual regression trees, with the best results being achieved using the techniques of bagging and random forest, as reflected in the measures of prediction error.

Kok et al. [17] examined the accuracy of decision tree (i.e. regression tree) model in comparison with the traditional hedonic pricing model (HPM) in predicting value and net operation oncome (NOI) of multi-family properties in several states in USA. They tested the two models using large datasets of over 28,145 unique properties. Results revealed the superiority of the decision tree model compared to HPM technique. The absolute error (MdAPE) between property value as predicted by the model and the actual transaction price is 9.3% for multifamily assets. Results showed that only 48% or less of property value is explained by property characteristics, whereas local amenities, census data, and local market dynamics contributed to the remaining 52%.

Dimopoulos and Bakas [18] used sensitivity analysis to investigate how random forest algorithm performed compared to other three algorithms (i.e. gradient boosting, linear regression, and non-linear regression) regarding predicting apartment prices. They found major similarity across the four methods as discussed in section 3.3 below.

Table 3. Applications of random forest and bagging algorithms in real estate. *

Study	Major findings	Domain of application
Ceh et al. [13]	The developed random forest model provided accurate prediction and significantly outperform traditional HPM approach, according to actual sales price and all performance measures.	Valuation of apartments & multi-family properties.
Aydinoglu et al. [14]	The random forest algorithm accurately predicted the value of residential land as evaluated by performance measure and IAAO standards.	Valuation of urban land (residential land).
Dimopoulos et al. [15]	The random forest random forest outperformed the linear regression models in predicting apartment's market value.	Valuation of apartments /multi-family properties.
Alfaro-Navarro et al. [16]	bagging and random forest performed better than boosting technique. Feasibility of applying the method to many cities through automating selecting best ML/DL model for particular city/district.	Value of apartments, automate the selection of best ML/DL model for each city of a country.
Kok et al. [17]	The developed decision tree model outperformed the HPM approach in predicting market value net operating income (NOI) of multi-family properties.	Valuation of multi-family properties, and the net operating income (NOI) of these properties.

* See also papers [18] & [24] that address these algorithms

3.3 Boosting and Gradient Boosting

Three papers have investigated gradient boosting algorithm, while another three papers have examined boosting algorithm as shown in table 4 and discussed below.

Dimopoulos and Bakas [18] empirically investigated the capabilities of four ML/DL algorithms (i.e gradient boosting, random forests, linear regression, and non-linear regression) and how they can be used for the mass appraisal of real estate properties. They used sensitivity analysis to examine how each algorithm perform regarding predicting apartment prices, and what property variables (property characteristics) are taken into account in predicting apartment price. Results demonstrated major similarity patterns across the four methods, in addition to some differences. The aim of the approach is to make the models more transparent for professional working in real estate field.

Vargas-Calderón & Camargo [19] investigated the ability of XGBoost, a gradient boosting model, to predict property prices based on data published by sellers on real estate market websites. The model which predicts whether a property has higher or lower price than the average price of its similar properties was evaluated based on a data set of 57,516 records. The experiment results demonstrated that the accuracy of a classifier that involves property descriptions by sellers is slightly higher than a classifier that only uses the features of the property. The authors recommend to add this model to features-based models to improve the accuracy of predicting property prices.

Gružauskas et al. [20] explored how to integrate ML algorithms for practicing real estate professionals where regulations require reasoning for value thus it is not allowed to mass valuation models to provide the final value of a real estate property. The authors proposed a model that uses the boosted decision tree algorithm as a supplement to the sales comparison method, a traditional real estate valuation method, when historical real estate transaction data is limited. The model predicts the average price of real estate in a region which can be used as correction coefficients in the sales comparison method where historical data is limited or unavailable.

As discussed in section 3.2, Alfaro-Navarro et al. [16] compared the performance of boosting algorithm with other two ensemble algorithms (bagging, random forest) and found that the bagging and random forest outperformed the boosting method. Gradient Boosting is also examined by Clark and Lomax [23] and Derdouri & Murayama [24] as detailed in section 3.6.

Table 4. Applications of boosting algorithms in real estate and major findings. *

Study	Major findings	Domain of application
Dimopoulos and Bakas [18]	major similarity patterns, in addition to some differences, between across the 4 ML/DL methods (gradient boosting, random forests, linear and non-linear regression).	Property valuation, mass appraisal.
Vargas-Calderón & Camargo [19]	that the accuracy of XGBoost algorithm working on property descriptions by sellers is slightly higher than a classifier that only uses the features of the property.	property valuation, predict property value.
Gružauskas et al. [20]	The proposed approached based on boosted decision tree is highly accurate, feasible and highly needed for practicing real estate professionals.	Single (object) property valuation.

* See also papers [16], [23] & [24] that address those algorithms

3.4 M5 Model Tree and Cubist

The Cubist model, which was developed by J.R Quinlan in 1992, is also referred to as M5 of which M5P is a special type. The M5 Model Tree / Cubist model is examined in 4 papers as shown in table 5 and discussed below. Bogdan et al. [21] conducted a comparative study using data of actual sales transactions to examine the performance and accuracy of three ML algorithms with two real estate expert algorithms in predicting the prices of residential buildings in urban areas in Poland. The ML algorithms are Pruned Model Tree (M5P), Multilayer Perceptron (MLP), and Linear Regression (LR), while the expert algorithms are: N-Nearest Similar Properties (NSP), N-Latest Transactions in an Area (LTA). Empirical results showed that ML algorithms surpassed the expert algorithms in accuracy. When comparing ML algorithms, the M5P revealed the best performance and linear regression showed the least accuracy. On the other hand, the LTA expert algorithm outperformed the NSP. Although the ML models performed better than the expert systems, the accuracy among them were not high: the median of MAPE, as a measure of accuracy in prediction, ranged from 10% to 13.4%.

Trawiński et al. [22] conducted an experimental study to evaluate the accuracy of predicting the prices of residential apartments by two ML algorithms (i.e. Pruned Model Tree -M5P, and Multilayer Perceptron -MLP) compared with three expert algorithms used in real estate valuation (N-Nearest Similar Transactions -NST, N-Latest Transactions in an Area -LTA, N-Random Similar Transactions -RST) using real-world data of actual real estate transactions. Each of M5P and MLP algorithms is applied as a base learner three times differently: single model, bagging ensemble, and additive regression. Empirical results revealed that bagging ensembles of both M5P decision trees and MLP gave better accuracy than single models and boosting ensembles. The ML also performed better than the expert systems. Nevertheless, all analyzed models and algorithms, but RST, could be employed in an automated system to support professional valuers in real estate appraisal.

Cubist algorithm is examined by Clark and Lomax [23] and Derdouri & Murayama [24] as detailed in section 3.6. In these two papers the cubist algorithm and its accuracy in predicting land and property value is investigated.

Table 5. Applications of M5 model tree / Cubist in real estate and major findings. *

Study	Major finding	Domain of application
Bogdan et al. [21]	ML/DL algorithms surpassed the expert algorithms in accuracy, though the difference is not high. For ML/DL models: M5P showed best performance while linear regression showed the least accuracy. The LTA expert algorithm outperformed the NSP	Valuation of sales prices of residential buildings.
Trawiński et al. [22]	Bagging ensembles of both M5P and MLP gave better accuracy than single models and boosting ensembles, and also outperformed the expert systems.	Valuation of apartment prices.

* See also papers [23] & [24] that address these algorithms

3.5 Support Vector Machine (SVM)

Four papers have studied the application of support vector machine (SVM) in real estate. Petkov [12] developed an application based on support vector machine (SVM) and self-organizing map (SOM) to predict mid-term rental prices based on various spatial variables in Lisbon city. He found that the ML methods (SVM and SOM) outperformed the non-ML methods (i.e. hedonic price model) and that the performance of ML methods depends on the type of variables used in the model.

Support vector machine (SVM) method is also examined by papers [23] and [24], as detailed in section 3.6 below. Derdouri & Murayama [24] conducted a comparative study to examine the accuracy of predicting land prices of 10 algorithms: 9 ML algorithms and 1 statistical method (the universal kriging regression). Two of those algorithms are SVM: support vector machines with linear kernel (SVMLinear) and Support vector machines with radial basis function kernel (SVMRadial) as detailed in section 3.6 below.

Table 6. Applications of SVM in real estate and major findings. *

Study	Major findings	Domain of application
Petkov [12]	ML methods (SVM and SOM) outperformed the non-ML methods (i.e. hedonic price model) in predicting rental prices. Performance of ML methods depends on type of used variables.	mid-term rental prices.
Derdouri & Murayama [24]	Tree-based models (cubist, random forest & gradient boosting) performed the best in estimating land prices, while linear regression models were the worst models. Random forest ranked the best. The universal kriging regression model ranked better than several ML models examined.	Valuation of land.

* See also papers [23] & [24] that address these algorithms

3.6 Linear and Non-Linear Regression Model

Regression models are the second most used ML methods in real estate, after the neural networks. As shown in Table 1 and Figures 1 and 2, seven research papers (out of 20 papers) have examined the application of 7 regression models in the context of real estate development and valuation. The regression algorithms and the papers that examined them are as follows: the generalized additive model using splines (paper [24]), generalized linear model (paper [24]), k-nearest neighbors (paper [24]), linear regression (papers [21], [15], [18]), logical regression model (paper [9]), and multivariate adaptive regression splines (paper [23], [24]), and non-linear regression (paper [18]).

Clark and Lomax [23] investigated the accuracy of predicting rental prices of properties using four ML algorithms: Cubist, gradient boosting, multivariate adaptive regression splines (MARS), and Support Vector Machine (SVM), in addition to one hedonic model traditionally used in valuating real estate properties which is generalized

linear model (GLM). The study used large dataset of rental prices extracted from property listings web site that covers whole England. The data set include a wide range of residential properties types including bungalow, detached, semi-detached, terraced, apartment among others. Experimental results showed that ML algorithms outperformed the GLM. In addition, the two tree-based methods (Gradient Boost and Cubist) outperformed the regression-based approaches of SVM, MARS and GLM. The cubist performed the best.

Derdouri & Murayama [24] conducted a comparative study to examine the accuracy of predicting land prices of ten algorithms: 9 ML algorithms and 1 statistical method (the universal kriging regression). The 9 ML methods are 3 linear regression models generalized linear model (GLM), generalized additive model using splines (GAMS), support vector machines with linear kernel (SVMLinear); 3 non-linear regression models multivariate adaptive regression spline (MARS), k-nearest neighbors (kNN), support vector machines with radial basis function kernel (SVMRadial); and 3 tree-based models Cubist, stochastic gradient boosting, random forest. The results revealed the superiority of the tree-based models (Cubist, random forest and gradient boosting) in estimating land prices accurately, while linear regression models were the worst models. Random forest was the most robust method among all methods in estimating land prices reliably. The universal kriging regression model ranked better than other ML models examined.

Table 7. Applications of regression models in real estate and major findings. *

Study	Major findings	Domain of application
Clark and Lomax [23]	The two tree-based methods (gradient boost and cubist) outperformed the regression-based approaches of SVM, MARS and GLM.	Valuation of rental prices of wide range of properties.
Derdouri & Murayama [24]	Tree-based models (cubist, random forest & gradient boosting) performed the best in estimating land prices, while linear regression models were the worst models. Random forest ranked the best. The universal kriging regression model ranked better than several ML models examined.	Valuation of land.

* See also papers [21], [15], [18] & [9] that address these algorithms

4 Results and Discussion

The results show that the analyzed research papers examined the use of 20 different types of ML/DI algorithms, as summarized in Figure 1. As illustrated in Figure 1, the analyzed 20 research papers have examined a total of 43 ML/DL algorithms, of which only 20 algorithms are unique ones, as listed in the left-side column of Figure 1. The number of ML/DL algorithms examined in each paper ranged between one to nine.

Used ML/DL algorithm		Paper/study																							
		5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24				
Neural Network	Artificial neural network	√	√	√	√																				
	Convolution Neural Network					√	√																		
	Recurrent neural network					√		√																	
	Multilayer perceptron																		√	√					
	Self-organizing map									√															
RE, B, DT	Bagging													√											
	Random forest										√	√	√	√		√									√
	Decision tree														√										
Boost	Boosting													√											
	Gradient boosting															√	√							√	√
Cub, M5	Cubist																						√	√	√
	Pruned model tree																					√	√		
Regression models	Support vector machine										√												√	√	2
	Linear regression													√			√					√			
	Generalized linear model																								√
	Generalized additive model using splines																								√
	Non-linear regression																√								
	Multivariate adaptive regression spline																							√	√
	K-nearest neighbors																								√
	Logical regression model						√																		
	Total # of used ML/DL algorithms	1	1	1	1	3	1	1	2	1	1	2	3	1	4	1	1	3	2	4	9				

Fig. 1. The ML/DL algorithms used in the 20 analyzed papers.

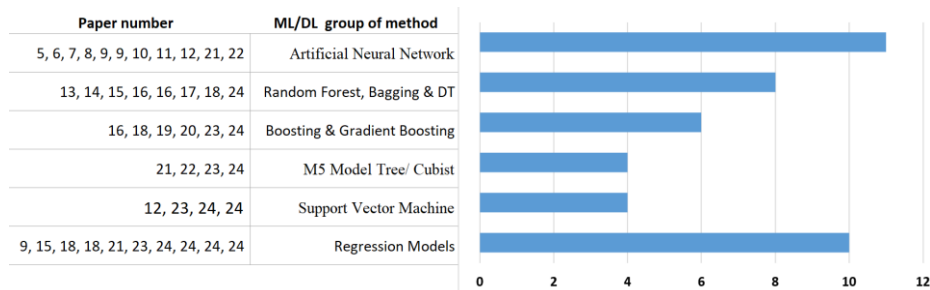


Fig. 2. Frequency of use of ML /DL algorithms in the examined research studies.

For further analysis, the 20 ML/DL algorithms/models are grouped in six thematic groups/categories (see Figure 2). Figure 2 shows the frequency of use of ML /DL algorithms across the six main groups of ML/DL algorithms. As shown in Figure 1, the neural network is the most used algorithm (examined 11 times in 10 papers), followed by regression models (examined 10 times in 6 papers). Tables 2 to 7 shows that the main application of ML/DL models is oriented toward valuation and the estimation of

market value and prices of various types of real estate such as land, apartments, single family house, villas and semi-detached villas as well as rental prices.

The scope of the analyzed papers falls under four categories: 1) compare the performance of ML with non-ML algorithms traditionally used in the real estate field, 2) compare the performance of different ML algorithms, 2) optimize ML algorithms for real estate through novel applications, and 4) general utilization and exploration of the use of ML/DL models in real estate field. Figure 3 provides an outline of the scope of the analyzed papers as related to these four categories. As shown in Figure 3, most of the papers (10 out of 20) are focused on comparing the performance of ML/DL algorithms with other ML/DL algorithms or with non-ML algorithms (category 1 and 2), while 7 papers investigated optimizing ML/DL algorithms for real estate use through novel applications (category 3), while 3 papers explored the general utilization and applicability of ML/DL in real estate (category 4).

Scope of the study	Paper/study																							
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24				
Compare performance of ML with non-ML algorithms			√	√			√	√	√				√				√	√	√	√				
Compare performance across ML Algorithms					√			√			√	√		√			√	√	√	√				
Optimize ML algorithms for real estate /novel application		√	√			√	√			√		√	√											
General utilization & exploration of ML applicability in RE.	√														√	√								

Fig. 3. The scope of the examined research papers as related to utilizing ML/DL algorithms in real estate.

5 Conclusions and Future Prospects

This paper provides a systematic survey of the application of ML and DL algorithms in real estate development and valuation. It provided a comprehensive analysis of ML and DL methods that have been applied and examined in real estate research in the past four years (2017-2020) as investigated in 20 empirical studies. It systematically described and characterized the ML/DL techniques used in the real estate field and their potential to advance and improve a wide range of activities and processes in real estate development and valuation.

The results show that there is high increase in using ML and DL algorithms in real estate valuation and appraisal during the past decade or so. The ML and DL methods have been deployed in several domains in real estate such as valuation of land, renting and market value of real estate properties. However, the application of ML and DL in this field still new and more growth is expected in the future.

Several types of classification were used to identify the most frequently used ML/DL models and the domains of application of these models in the real estate field. The results showed that the most used ML/DL algorithms in real estate were neural networks (5 variations of the algorithm, examined in 11 studies), regression models (7

variations of the model, examined in 7 studies), random forest (used in 6 studies), booting (2 variations of the algorithm, examined in 6 studies), cubist / M5P (used in 4 studies), support vector machine (2 variations of the model, used in 3 studies). These studies have examined the prediction power and accuracy of the ML/DL models particularly in property and land valuation, and their potential to improve data quality and interpretability and advance work in the real estate field. They also highlighted the challenging problems encountered and presented possible solutions to these challenges.

Acknowledgment

The author acknowledges the support of the King Fahd University of Petroleum and Minerals for the research project.

References

1. Rizun, A., Baj-rogowska, A.: Can Web Search Queries Predict Prices Change on the Real Estate Market. *IEEE Access* 9, 70095 – 70117 (2021).
2. Jahanshiri, E., Buyong, T., Shariff, A.R.M.: A review of property mass valuation models. *Pertanika Journal of Science & Technology* 19 (1), 23-30 (2011).
3. d'Amato, M., Kauko, T.: *Advances in automated valuation modeling: AVM after the non-agency mortgage crisis*. 1st edn, Springer (2017).
4. Zhou, G., Ji, Y., Chen, X., Zhang, F.: Artificial Neural Networks and the Mass Appraisal of Real Estate. *International Journal of Online and Biomedical Engineering (iJOE)* 14(3), 180-187 (2018).
5. Abidoeye, R.B., Chan, A.P.C.: Modelling property values in Nigeria using artificial neural network. *Journal of Property Research* 34(1), 36-53 (2017).
6. Guo, Y., Lin, S., Ma, X., Bal, J., Li, C.: Homogeneous Feature Transfer and Heterogeneous Location Fine-tuning for Cross-City Property Appraisal Framework. In Zhao, Y., Warwick, G., Stirling, D. Li, C., Koh, Y.S., Islam, R., Islam, Z. (eds.) *DATA MINING - 16TH AUSTRALASIAN CONFERENCE (AusDM 2018)*, pp. 161-174. Springer, London (2019).
7. Alexandridis, A. K., Karlis, D., Papastamos, D., Andritsos, D.: Real Estate valuation and forecasting in non-homogeneous markets: A case study in Greece during the financial crisis. *Journal of the Operational Research Society* 70(10), 1769-1783 (2019).
8. Abidoeye, R.B., Chan, A.P.C.: Improving property valuation accuracy: a comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal* 24(1), 71-83 (2018).
9. Yu, L, Jiao, C., Xin, H., Wang, Y., Wang, K.: Prediction on Housing Price Based on Deep Learning. *International Journal of Computer & Information Engineering* 12(2), 90-99 (2018).
10. Poursaeed, O., Matera, T., & Belongie, S.: Vision-based real estate price estimation. *Machine Vision and Applications* 29, 667–676 (2018).
11. Bin, J., Tang, S., Liu, Y., Wang, G., Gardiner, B., Li, Z., Li E.: Regression model for appraisal of real estate using recurrent neural network and boosting tree. In *THE 2017 2ND IEEE INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND APPLICATIONS (ICCIA)*, pp. 209-213. IEEE, Beijing, China, (2017).

12. Petkov, M.: Evaluation of Spatial Data's Impact in Mid-Term Room Rent Price Through Application of Spatial Econometrics and Machine Learning: Lisbon. MS thesis, Geospatial Technologies, University NOVA de Lisboa, Lisbon, Portugal, (2020).
13. Ceh, M., Kilibarda, M., Lisec, A., Bajat, B.: Estimating the performance of random forest versus multiple regression for predicting prices of apartments. *ISPRS International Journal of Geo-Information* 7(5), 168 (2018).
14. Aydinoglu, A. C., Bovkir, R., Colkesen, I.: Implementing a mass valuation application on interoperable land valuation data model designed as an extension of the national GDI. *Survey Review* 53(375), 349-365 (2020).
15. Dimopoulos, T., Tyrallis, H., Bakas, N.P., Hadjimitsis, D.: Accuracy measurement of Random Forests and Linear Regression for mass appraisal models that estimate the prices of residential apartments in Nicosia, Cyprus. *Advances in Geoscience* 45, 377–382 (2018).
16. Alfaro-Navarro, J., Cano, E.L., Alfaro-Cortés, E., García, N., Alfaro-Cortés, M., Larraz, B.: A Fully Automated Adjustment of Ensemble Methods in Machine Learning for Modeling Complex Real Estate Systems. *Complexity*, 2020, 1-12 (2020).
17. Kok, N., Koponen, E., Martínez-Barbosa, C.A.: Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal of Portfolio Management* 43(6), 202-211 (2017).
18. Dimopoulos, T., Bakas, N.: Sensitivity Analysis of Machine Learning Models for the Mass Appraisal of Real Estate: Case Study of Residential Units in Nicosia, Cyprus. *Remote Sensing*, 11, 3047, (2019).
19. Vargas-Calderón, V., Camargo, J. E.: A model for predicting price polarity of real estate properties using information of real estate market websites, (2019).
20. Gružasuskas, V., Kriščiūnas, A., Čalnerytė, D., Navickas, V.: Analytical method for correction coefficient determination for applying comparative method for real estate valuation. *Real Estate Management and Valuation* 28(2), 52-62 (2020).
21. Trawiński, B., Telec, Z., Krasnoborski, J., Piwowarczyk, M., Talaga, M., Lasota, T., Sawiłow, E.: Comparison of Expert Algorithms with Machine Learning Models for Real Estate Appraisal. In: Jędrzejowicz, P., Yildirim, T., Czarnowski, I. (eds.) *IEEE INTERNATIONAL CONFERENCE ON INNOVATIONS IN INTELLIGENT SYSTEMS AND APPLICATIONS INISTA 2017*, p. 51-54. IEEE, Gdynia, Poland (2017).
22. Trawiński, B., Lasota, T., Kempa, O., Telec, Z., Kutrzyński, M.: Comparison of Ensemble Learning Models with Expert Algorithms Designed for a Property Valuation System. In: Nguyen, N.T., Papadopoulos, G.A., Jędrzejowicz, P., Trawiński, B., Vossen, G. (eds.) *COMPUTATIONAL COLLECTIVE INTELLIGENCE 9TH INTERNATIONAL ICCCI 2017*, pp. 317-327. Springer (2017).
23. Clark, S. D., Lomax, N.: A mass-market appraisal of the English housing rental market using a diverse range of modelling techniques. *Journal of Big Data* 5(1), 43 (2018).
24. Derdouri, A., Murayama, Y.: A comparative study of land price estimation and mapping using regression kriging and machine learning algorithms across Fukushima prefecture, Japan. *Journal of geographical Sciences* 30, 794–822 (2020).