IPTI White Paper

Prepared for the Property Valuation Services Corporation, Nova Scotia, Canada

The Potential of Artificial Intelligence in Property Assessment

July 2021

International Property Tax Institute
5 Kodiak Crescent, Unit 10
Toronto, Ontario, Canada M3J 3E5
www.ipti.org

About the International Property Tax Institute

The International Property Tax Institute (IPTI) is widely recognized as the world's leading organization on property tax policy and practice.

IPTI's mission is to provide impartial, objective expert advice in the area of property tax systems and promote the concept that these systems should be fair and equitable and meet the needs of all stakeholders, i.e., governments, taxpayers, practitioners and academics. In addition, IPTI seeks to ensure that property tax systems contribute to the provision of high-quality services for the benefit of communities.

IPTI is a not-for-profit organization comprised of experts who support stakeholders in developing and maintaining effective and efficient property tax systems by providing them with:

- Research and analytical information
- Impartial, objective policy advice
- Strategic advisory and consulting services to create, test and implement policy, and to improve performance through innovative good practice
- Education and training services to enhance professional development and build technical competence
- Property information services to enable more effective decisions

In addition, IPTI specializes in:

- Property valuation processes: including data collection, mapping and data management; mass appraisal valuation for residential and non-residential properties; quality control
- Property tax collection and enforcement
- Appeal systems
- Technology and process integration and implementation, including data management, data analysis and reporting systems
- Electronic and on-line learning
- Sharing best practice

IPTI has a Board of Advisors which is comprised of internationally respected professionals all of whom have extensive experience in their respective fields. The breadth of membership of the Board reflects IPTI's commitment to international participation and sharing best practice on a global basis. The Board contributes to the strategic direction and overall planning for IPTI.

More information about IPTI can be found on its website www.ipti.org

International Property Tax Institute

Contents

Acknowledgements

IPTI wishes to acknowledge the support of the Property Valuation Services Corporation (PVSC) who commissioned this paper in order to advance understanding of artificial intelligence and the role it can play in the modern assessment office.

In particular, we would like to thank the following individuals for sharing their vision, knowledge, and experience with us (positions may have subsequently changed):

- **Kathy Gillis**, Chief Executive Officer
- **Jeff Caddell**, Internal Audit and Compliance Manager
- **Ashley Wu**, PhD, Chief Data Scientist
- **Dr. Hong Gu**, Professor of Statistics, Dalhousie University

The IPTI team working on this project consists of:

- Robert Gloudemans, Mass Appraisal Consultant
- Paul Sanderson, President, IPTI

Robert Gloudemans is an independent mass appraisal consultant. He previously worked for the IAAO and Arizona Department of Revenue and has served over 100 clients internationally. He is a prolific contributor to IAAO standards, courses, and publications.

Paul Sanderson is President of IPTI. Prior to joining IPTI, Paul had over 40 years' experience at the largest valuation agency in the world, the UK Valuation Office Agency, where he was the Director of Professional Services and responsible for all aspects of the valuation work undertaken by the agency.

IPTI would also like to thank the following members of a panel of experts who kindly let us have their comments on the draft version of this paper (positions may have subsequently changed):

- **Paul Bidanset**, Valuation Research Project Manager, International Association of Assessing Officers (IAAO)
- **Richard Borst**, Senior Research Scientist, Tyler Technologies, USA
- **James Ellens**, Data Science Manager, Municipal Property Assessment Corporation (MPAC), Canada
- **Brian Guerin**, Director, Assessment Standards & Mass Appraisal, Municipal Property Assessment Corporation (MPAC), Canada
- **Marco Kuijper**, Netherlands Council for Real Estate Assessment, The Netherlands
- **William J McCluskey**, University of Pretoria, South Africa
- **Ron Rakow**, Pleasant Lake Advisors, USA

The opinions expressed in this document are entirely IPTI's and it should be noted that any shortcomings or errors are also IPTI's alone.

Executive Summary

1. For the purposes of this white paper, and in the context of property valuation, IPTI defines Artificial Intelligence (AI) as machine learning (ML) designed to predict an outcome or provide an estimate of value, e.g., most probable sale price.

2. The vision of this white paper is to provide a first step toward the production of a standard on the use of AI in property assessment administration. While existing mass appraisal tools can be highly effective and produce excellent performance results, AI offers another viable tool that, if used properly, can efficiently produce equally and, arguably, more accurate valuations for many jurisdictions.

3. For that reason, IPTI considers the time has come for serious consideration of AI by the assessment community and hopes to see the guidance offered in this paper adopted on a more formal basis, either as a standalone standard or as an addition to an existing standard.

4. In the world of property valuation, AI has been extensively adopted in the private sector and interest, and its use, has now emerged in the public sector as well. As AI becomes more accepted and entrenched into our daily lives, stakeholders may be increasingly open to government sector uses.

5. This white paper presents a high-level summary of some of the most common AI methods, including model-based methods such as "expert systems" and adaptive learning methods, including Classification and Regression Trees (CART), Random Forests, Neural Networks, and Gradient Boosting Methods (GBM).

6. AI may offer a number of advantages including:

   - Improved accuracy
   - Greater objectivity
   - Enhanced speed
   - Reduced cost

7. AI also brings a number of potential issues including:

   - Challenges in "explainability"
   - The need for advanced technical expertise
   - Limited appraiser interaction and direction
   - Potential for over-fitting

8. This white paper identifies a number of technical factors that need to be addressed. They include:

    - Data preparation
    - Sample size and sufficiency
    - The need to account for date of sale
    - Parameter dependency
    - Quality assurance/control
    - Deployment

9. IPTI concludes:

    - With the appropriate expertise and deployment strategy, AI may produce values that are more accurate than those obtained by traditional equation-based approaches. Additional advantages include objectivity, efficiency, and cost.
    - AI methods are widely used in private sector appraisals and are likely to get increased consideration, testing, and potential use in the public sector, either as a primary or supplemental valuation method.
    - Issues associated with AI include the lack of a tangible equation, user understanding and acceptance, and CAMA system interface.
    - Because of AI's susceptibility to over-fitting, use of a holdout group or cross-validation is essential.
    - As in other appraisal methods, success depends on data sufficiency and integrity. In addition, a jurisdiction must possess the necessary data analyst(s) and supporting software.
    - AI-based values need not be standalone values but can be compared against, and used as a check on, more traditional values. Once an assessor has gained confidence with AI and addressed relevant issues, it can be given more prominence or used as the primary valuation method.
    - AI may not be for everyone. While some jurisdictions will find that the benefits of AI offset its drawbacks, others are producing accurate and equitable values using current methods, lack the required technical expertise, or simply prefer their current system. In short, AI provides a powerful option for those who have the disposition to use it and the data scientists to test and potentially adopt it.
    - IPTI recommends that the assessment industry consider and debate the guidance offered in this paper as part of a process to have AI adopted on a more formal basis, either as a standalone standard or as an addition to an existing standard.
    - Assessors can learn much from AI applications for private sector appraisals, as well as from the experience of jurisdictions pioneering and testing the use of AI methods.
    - Particularly helpful, however, would be a peer-approved standard or guide on the development and deployment of a credible, properly tested, and defensible AI-based appraisal system.

- The development of such a standard is a key objective of this white paper and IPTI recommends that the assessment profession define a concrete strategy for producing an AI standard that will guide assessors who are interested in exploring - and potentially adopting - this promising technology.

10. In addition, IPTI offers the following guidance for successful adoption of AI in property assessment:

- A jurisdiction must have the necessary management commitment and staff buy-in. Changes in the organization's structure and job responsibilities may be required.
- A jurisdiction must have, or contract for, the necessary technical expertise and enabling software.
- Data integrity and preparation are paramount.
- Determination of optimal methods and parameters requires testing and analysis.
- Steps should be taken to avoid over-fitting; use of a holdout group or cross-validation is essential.
- Attention should be given to such technical issues as outliers, time-of-sale, seldom-occurring features, and atypical properties.
- Predicted values should be scrutinized through traditional ratio studies and procedures should be adopted for handling problems or inequities that are found.
- Procedures should be in place for explanation of the chosen valuation approach, demonstrating equity, and evaluating and supporting values through comparable sales and/or other accepted procedures.
- Potential users should explore and learn from the experience of other jurisdictions who are using, or have tested the use of, AI.
- Jurisdictions need not take a "leapfrog" approach. They can use AI as a supplementary method and optionally increase reliance on it once they have gained experience and confidence in it and addressed related appraisal and change management issues.

1. Introduction and Purpose of Paper

What is artificial intelligence? Although many definitions are available, in the context of property valuation, IPTI defines artificial intelligence (AI) as *machine learning designed to predict an outcome or provide an estimate, e.g., most probable sale price.* AI, at least in IPTI's use of the term, is based on pattern and image recognition. It has the ability to process large volumes of data and requires intensive computer power of the type available in today's higher-end PCs and cloud services.

For purposes of this white paper, AI does not include standard statistical algorithms - most prominently multiple regression analysis (MRA) - in which the user specifies and calibrates a prediction model. Although users specify the dependent and independent variables, AI models produce no tangible equation.

This white paper is therefore distinct from the International Association of Assessing Officers (IAAO) standards on Mass Appraisal of Real Property and Automated Valuation Models which focus on equation-based applications of the three approaches to value in mass appraisal.

The vision of this white paper is to provide a framework or first step toward the production of a standard on the use of AI in property assessment administration. While existing mass appraisal tools can be highly effective and produce excellent performance results, AI offers another viable tool that, if used properly, can efficiently produce equally and, arguably, more accurate valuations for many jurisdictions.

Therefore, IPTI considers the time has come for serious consideration of AI by the assessment community and we hope to see the guidance offered in this paper considered and debated as part of a process to have AI adopted on a more formal basis, either as a standalone standard or as an addition to an existing standard.

2. Current State of the Art

AI has been, or is being, adopted into ever-expanding aspects of our daily lives ranging from targeted advertising and social media to self-driving cars and medical diagnosis. So many companies are now using AI that there are over 20 exchange traded funds (ETFs) with baskets of them.

The OECD has developed a set of *Principles on AI* designed to promote beneficial use, trustworthiness, safety, security, and disclosure [OECD 2019]. Adopted in May 2019, these principles have been signed by member countries and many non-member countries. Shortly after, in June 2019, the G20 nations adopted a similar set of AI principles.

In the world of property valuation, AI has been extensively adopted in the private sector and interest and use has emerged in the public sector as well.

2.1 Private Sector Valuation

The private sector has embraced AI for a variety of property valuation and related uses including mortgage approvals or pre-qualifications, refinancing, home equity loans, underwriting, site analysis, acquisition and divestiture decisions, and portfolio evaluation. Vendors obtain data from various sources beginning with public records obtained from assessors and recorders, which are increasingly being supplemented by sources ranging from macro-economic to "hyper-local" economic and demographic data. The latter can include such factors as distance to public transportation and coffee shops and the popularity of local restaurants. With many vendors to choose from, each claiming product superiority, some users opt for a "cascade" approach in which they average or blend several value estimates. One vendor even evaluates whether AI would be suitable for a particular property and, if so, recommends appropriate ones to clients.

While residential valuation for mortgage-related purposes was one of the first uses, more recent applications include online buy, sell, and refinance offers. Vendors are beginning to use imagery technology based, for example, on scanned photos used in listings to assess subjective features like construction grade, condition, and buyer appeal.

Although later to the table, applications to commercial properties are also burgeoning with some AI vendors specializing in commercial valuation and location analysis. In addition to public records, data comes from REITs, government regulatory agencies, appraisal reports, existing clients, open street maps, and satellite imagery. Another application is determining rent potential and fair market rents for existing properties.

As in property assessment, private sector performance metrics are based on a comparison of predicted and actual prices and include R-Square, the standard deviation, and the mean absolute percent error (akin to the COD). Another widely used measure is the percentage of errors that fall within a stipulated range, most typically 10%. In addition, private sector

AVMs commonly provide a forecast standard deviation (FSD) for each value estimate, the calculation of which is proprietary and varies among companies and in reliability.

Professional valuation organizations, including the Appraisal Foundation (AF), the Royal Institution of Chartered Surveyors (RICS), the Appraisal Institute of Canada (AIC), the Appraisal Institute (AI), the International Valuation Standards Committee (IVSC), the European Group of Valuers Association (TEGoVA), the Australian Property Institute (API), and the World Association of Valuation Organisations (WAVO), have all addressed the use of AVMs to some extent.

In general, members are permitted to use AVMs as long as they have a rudimentary understanding of the methodology, gauge the AVM to be appropriate to the assignment, are confident that it utilizes the appropriate data, and understand - and can convey - the generated output. Both USPAP in the US and CUSPAP in Canada do not regard an AVM output as an appraisal, but allow that it can become the "basis" for a credible appraisal if the appraiser performs due diligence in a review of the value and transmission to the client in what is sometimes referred to as an appraiser-assisted AVM.

2.2 Public Sector Property Assessment

In contrast to the private sector, prior to 2020, AI had not been used for property tax assessment purposes. The reasons are varied, but among the most important, are the lack of competition that has fuelled private sector applications, the demand for transparency in assessments, and the complexity of AI.

However, change is on the horizon as AI becomes more widely accepted, its benefits demonstrated, and pioneering jurisdictions take the lead and share their experiences in its use.

Perhaps the most notable of these is the Property Valuation Services Corporation (PVSC) of Nova Scotia, Canada, which is using AI to produce values for a significant proportion of its residential parcels. PVSC has made several presentations on its research and results at IAAO and IPTI conferences. As part of its due diligence, PVSC commissioned IPTI to review its methodology, plans, and preliminary values. While making a number of technical recommendations, the relevant IPTI report states in its conclusions that *"PVSC has done its homework and developed a process that stands to produce more accurate values than traditional mass appraisal approaches. IPTI also considers that PVSC has the necessary staff and wherewithal to build those models."*

Another large, rapidly growing US jurisdiction (Wake County, North Carolina) has contracted with a software company to produce AI values to use as a check on its inhouse values, as well as to suggest similar neighbourhoods that can be combined for appraisal analyses and comparables selection. A number of other jurisdictions have researched and tested AI for residential valuation with favourable results.

AI can also be leveraged for other assessment applications than valuation alone. Many jurisdictions, for example, are using AI for aerial change detection and several Louisiana parishes are using it in the audit of homestead exemptions.

In any case, while applications of AI to public sector property valuation have, to date, been limited, a number of jurisdictions are either using or researching the method. Given its success and acceptance in the private sector, the potential for expanded use of AI for property assessment must be considered.

3. AI Methods

This section presents a high-level summary of some of the most common AI methods. Just as with MRA, there are various versions of the conceptual methods described here.

3.1 Model-Based Methods

To a large extent, model-based methods were a forerunner to the adaptive learning methods described in 3.2 below.

Some model-based methods take the form of "expert systems" designed to replicate the modelling steps and decisions that an experienced expert might make, thus eliminating the need for the otherwise requisite modelling skills and experience. Outside of the appraisal world, perhaps the best known such system was Deep Blue, the IBM supercomputer programmed to beat world chess champion Gary Kasparov in a famed 1997 chess match by optimally countering his every possible move.

In the modelling world, REX (Regression Expert) developed at Bell Labs and THESA (The Expert Statistical Assistant) developed by Boeing Software Services are early examples of rule-based software programmed to conduct statistical analyses in much the same manner as an expert might. A 1990 article by Jensen in the IAAO Property Tax Journal describes an expert system for CAMA developed by Siemens and Sigma Systems Technology. Like REX and THESA, the system had embedded rules and analytics for data screening and cleansing, transforming variables, testing MRA assumptions, generating a model, and reporting performance statistics. Such systems never found traction in the assessment community because of their complexity, proprietary nature, functional limitations, difficulty in explaining output, and deployment issues.

Somewhat akin, however, are comparable sales algorithms that are programmed to select the most comparable sales for a subject property and sometimes adjust for differences between comparables and the subject property. Of course, in the better of such systems, the user can intervene to modify the selected comparables and adjustments.

Some advanced variations of MRA utilize pattern recognition or expert system techniques. The most extreme of these is Best Subset Regression which identifies the optimal combination of candidate variables through a computer-intensive iterative process. Ridge regression and Lasso (least absolute shrinkage and selection operator) regression are techniques to reduce multicollinearity effects and avoid overfitting that might occur in a purely least squares solution. Ensemble regression is a multi-stage process that focuses on reducing errors from the prior model. Perhaps most interesting are generalized additive models (GAM) in which continuous variables are internally smoothed or transformed to provide a better linear fit and thus predictive accuracy. Again, however, because of their complexity alone (not to mention interface with a CAMA system), none of these techniques are used, or likely to gain favour, in property assessment.

3.2 Adaptive Learning Methods

Much more promising are adaptive learning methods that, while they do not produce a valuation equation, are widely used in the private sector and have demonstrated their ability to achieve predictive accuracy generally superior to equation-based methods.

- Classification and Regression Trees (CART). The most straight-forward and understandable of these methods, CART iteratively splits and re-splits the data so as to minimize the resulting variance (squared errors) of each resulting "leaf" or node. Each split determines the optimal variable to split on and the optimal split point for that variable. The process continues until a pre-set stopping point is reached, which is often the minimal number of data points (sales) required for each leaf. Estimated values are based on a sample average (or other function) of the sales prices in each leaf. Prediction accuracy may be improved (and over-fitting minimized) by testing the resulting tree on a test dataset and pruning branches to improve accuracy on the test set.

- Random Forests. Random forests are an ensemble - or "forest" - of decision trees in which, to avoid over-fitting, each tree is based on a random subset of the data. Additionally, each split of a tree can be based on a random subset of the available variables. Predictions are based on a combination (e.g., simple average) of those produced by each tree in which an observation occurs. Again to avoid over-fitting, random forests are typically "shallow" trees. User-determined parameters include the number of trees, percentage of data to use to build each tree, number of features (variables) considered in each split, and minimum leaf size. As in CART, the analyst can set aside a holdout group.

  As can be seen, random forests were designed to minimize over-fitting of data. The relative importance or predictive power of the variables can be evaluated by determining how much they decrease variance when used in splits. Although simple in concept, unlike CART, it is impossible to demonstrate how individual predictions were determined by following the splits or branches of a single tree.

- Neural Networks. Neural networks might be described as the ultimate black box. The user specifies an input layer (variables) and the software develops an output layer (predictions) based on hidden layers that uncover patterns and relationships in the data and develop weights for each input variable. Various software products offer different variations on the method. Users can specify a holdout group and such parameters as number of hidden layers (one is often sufficient), the number of nodes in a layer used to join and weight the inputs, the number of iterations, and learning rate. The software may be able to indicate the implicit importance of each variable by showing, graphically, how predictions change as a variable is changed

with all others held to their averages, or by comparing average errors or R-square values with and without the variable included.

- Gradient Boosting Methods (GBM). GBM appears to be emerging as the AI software of choice for property valuation. It is a branching approach in which sales divide or branch along different paths to form combinations of property features associated with similar prices. After forming an initial "tree", errors are analysed, and a second tree built to reduce residuals (errors) from the first tree. Subsequent trees are built in the same manner, each based on errors from the prior tree. During the process, observations are weighted based on their remaining errors, so that the model increasingly strives to fit the most problematic observations. The process continues until a stipulated number of trees has been built. Predicted values are calculated as the weighted sum of values from the first tree and refinements from subsequent trees. Properties that share the same leaf on a given tree are assigned the same adjustment for that tree. Thus, the predicted value for a property is a weighted sum of adjustments obtained from each of the leaves in the many trees in which the property appears.

  To minimize over-fitting, like random forests, GBM is often built on shallow trees with, say, three to six splits per tree. The method gains its strength from the increased refinements made during subsequent trees, each striving to reduce remaining errors from prior trees. It is a "slow learner" in that it adds only a small proportion of the effect of each newly fitted tree. This prevents the method from over-fitting the data too quickly, instead making small incremental adjustments as the relative contribution of the various candidate predictors is determined.

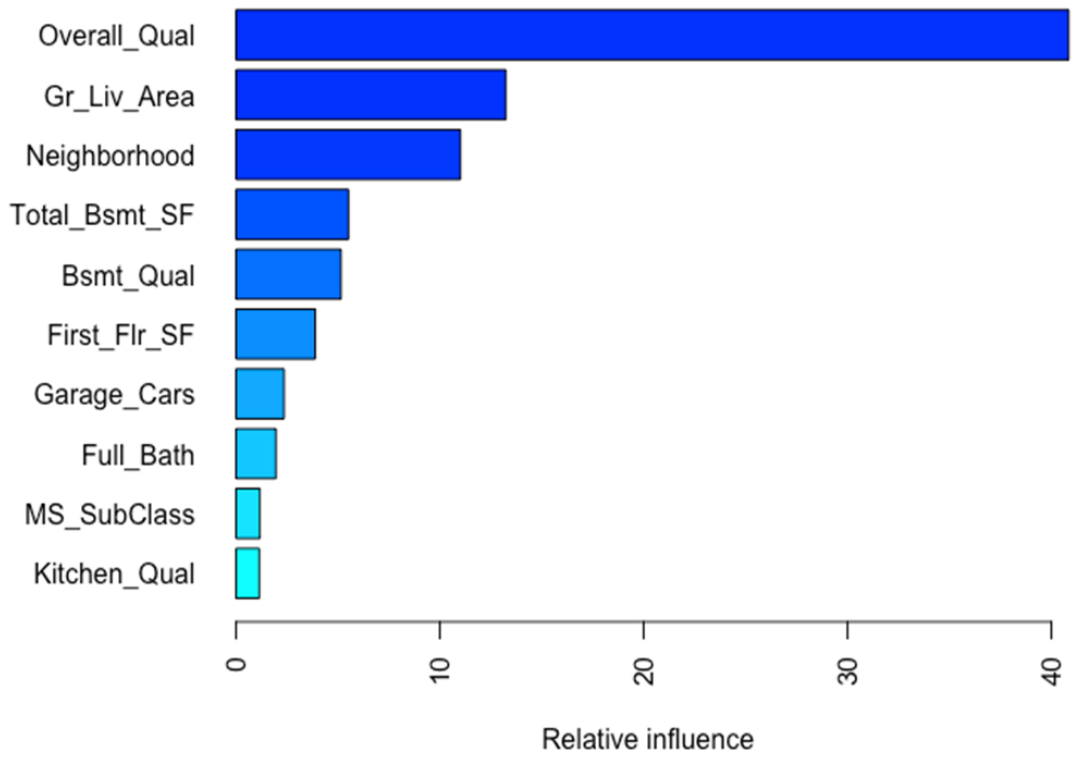
  As with other methods, holdout groups and cross-validation (in which models are developed and tested on different subsets of the data) can be used to better gauge expected performance for unsold properties. In addition, although again there is no model or tangible coefficients, the relative importance of variables can be illustrated by value influence plots, which indicate the relative improvement in the mean squared error when the variable is introduced in a tree, and partial dependence plots, which indicate the average change in the dependent variable as one variable is changed with the others held constant (see Figure 1 - next page).

  After extensive testing, PVSC concluded that GBM offered the best predictive performance among a number of AI methods tested. Others who have evaluated AI for property valuation purposes also appear to be gravitating toward it and/or neural networks.

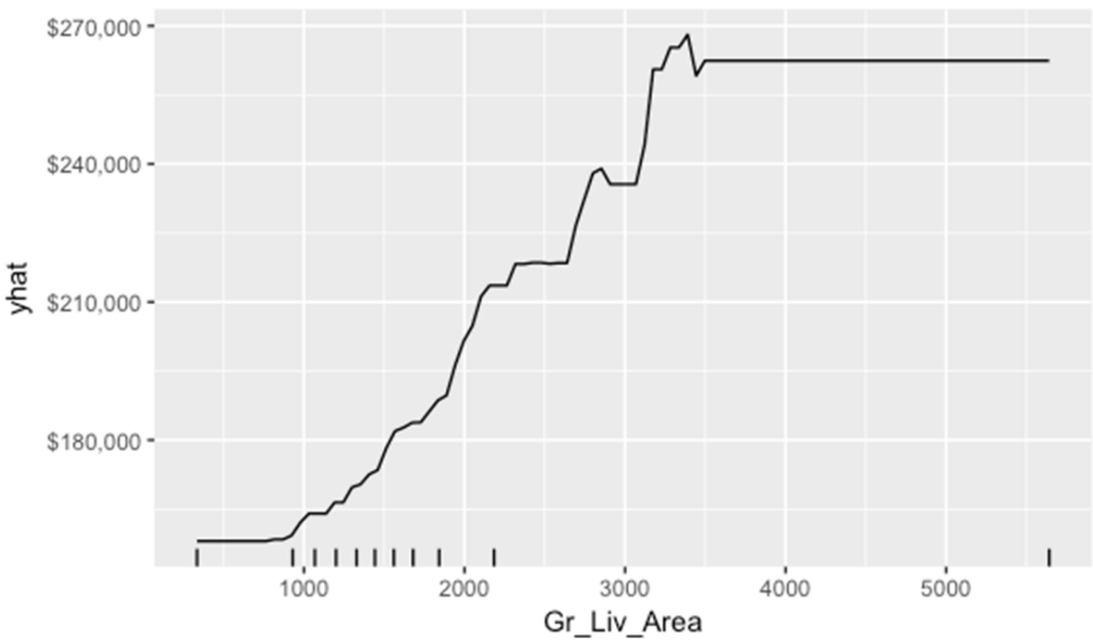
  Versions of AI methods are available in a number of statistical software packages including R, Python, SAS, and SPSS. In most software, users can determine sampling percentage, tree depth, number of trees, learning rate, and other parameters.

Figure 1 – Value Influence and Partial Dependency Plots

Example of Value Influence Plot



Example of Partial Dependence Plot



Source: UC Business Analytics R Programming Guide (June 2018)

3.3 Explaining AI Predictions

As illustrated by value influence and partial dependence plots, effective tools exist to illustrate the relative importance and contributions of individual variables in machine-learning models. Illustrating what determines a particular predicted value, however, is more difficult. One approach to doing so involves Local Interpretable Model-Agnostic (LIME) methods. The idea underlying LIME is that any predicted value, regardless of method, can be closely replicated by a simple surrogate (e.g., MRA) model in the vicinity of the property of interest (think subject property). Assume, for example, that we develop an AI model for a market area and that it predicts a value of $482,776 for a subject property. The model may return a low average error and we are comfortable with the relative contributions of the individual variables to the predictions, but how do we explain how the $482,776 was determined?

The idea behind LIME is that we can "perturb" the subject's data slightly, randomly varying living area and other features about their actual values, generate AI-values for the perturbed data set, and run a linear MRA (or other simple, explainable) model on the perturbed data with the dependent variable set to the predicted AI values. Of course, some variables, particularly lot size and age, may be far from linear. However, in the vicinity of the subject (e.g., in a tight range on these variables), the contribution can be taken as linear, much as one might approximate the effects of sale date through spline variables. LIME postulates that in the vicinity of the subject, the simple surrogate model will return a value substantially the same as the underlying AI model and thus one can use the surrogate model to explain the value of the subject property.

At least so the theory goes. A problem with property data is that many variables are discrete (not continuous) and thus cannot be entered directly into an MRA model regardless of its structure (additive, multiplicative, or hybrid). In addition, determining proper parameters for perturbing the data can be daunting and a separate model would have to be generated for each subject property of interest. While appealing in theory, these and other complexities bode ill for application of LIME to AI-based valuation models.

A more promising approach relates to "break-down" plots and Shapley Additive Explanations (Shapley values). The idea behind break-down plots is that one can see the impact of each variable on the predicted value by shifting the variable with the others held constant. The concept thus is not unlike a "base home approach" that computes the value impact of each difference in characteristics between the base (typical) property and subject property.

Of course, when variables behave interactively (as living area and quality generally do), the interactive portion of their impact is attributed to the variable we change first; generally living area in the base home approach. One option in break-down models is to order the variables according to their influence in the model.

As an illustration. Consider Figure 2 in which the mean predicted value for properties in a model is $511,629. The subject property is larger and changing the living area of each property to match that of the subject generates a mean value of $664,832. However, the subject is below average with respect to Quality, EffAge, and NBHD and changing those features to match the subject generates an average value of $532,226. Ultimately, when all data is changed to match the subject, we arrive at the value generated by the model itself, $468,988.

This is but one illustration. Depending on software employed, the plot might show the average change (rather than average predicted value) at each step or a visualization of the distribution of the predicted values or changes for all properties at each step.

Figure 2 – Example of Break-Down Plot

MeanESP			• 511,629			
LivArea				• 664,832		
Quality				• 638,534		
EffAge			• 584,757			
NBHD			• 532,226			
LotSize			• 551,962			
Garage			• 578,060			
FinBsmt			• 510,121			
Style		• 491,563				
Traffic		• 468,988				
	300,000	400,000	500,000	600,000	700,000	800,000

A limitation of break-down plots, of course, is that interactive effects are attributed to variables in the order they are considered. In Figure 2, had we considered NBHD before LivArea, we would undoubtedly get a different average predicted value and corresponding value change than $532,226. Although more complex and computer-intensive, Shapley values address this issue. The idea is that we can randomly perturb the order of the variables and determine the average contribution of each over the random orderings.

To generate Shapley values for a given variable for a given subject property, we specify a number of iterations. For each iteration we draw a random property other than the subject and order the variables randomly as well. The procedure then generates two comparison properties. The first contains data for the subject property up to and including the variable of interest and data for the random property thereafter. The second comparison property is the same as the first except that the variable of interest is obtained from the randomly drawn property.

Assume, for example, that we are interested in the impact of the absence of a garage on the value of a subject property. Specifically, we want to answer the same question as in a break-down analysis: how much will the average value change because our subject property does not have a garage?

Assume the nine variables in column 1 of Figure 3. Column 2 contains values for the subject property and column 3 values for a randomly drawn comparison property. The variables are randomly ordered from Quality to FinBsmt. Note that our subject property has no garage and the comparison property has a 2-car garage. Hypothetical property X1 contains the value of the subject for the first six variables (through Garage) and values for LotSize, Traffic, and FinBsmt from the comparison property. Property X2 has the same features as X1 except that it takes the value of Garage from the comparison property.

Figure 3. Example of Shapley Calculations

	Subject	Random	X1	X2
Quality	3.5	4	3.5	3.5
Lake	0	0	0	0
EffAge	53	32	53	53
LivArea	1616	1842	1616	1616
NBHD	30102	30109	30102	30102
Garage	0	2	0	2
LotSize	5602	8441	8441	8441
Traffic	1	0	0	0
FinBsmt	658	0	0	0
Predicted			316,445	329,820
Diff (X1-X2)			-23,375	

As shown in Figure 3, the difference in value is -$23,375. That is, based on the comparison, absence of a garage vs having a 2-car garage when other features are equal, reduces value by $23,375. We repeat the analysis many times drawing properties with various garage counts and other features and obtaining various value differences. The Shapley value is the average of the differences (X1-X2) and, with many iterations, the absence of a garage is, in the end, compared against the conglomerate distribution of garages in the data set. Hence, it represents the expected change in average value due to absence of a garage.

The reason for randomly ordering the variables and taking values for variables through the variable of interest from the subject property and values for subsequent variables from the randomly drawn comparison property is to spread interactive effects randomly and evenly across the variables.

The relative effect of multiple characteristics on the value of a subject property can be illustrated through "force" plots as illustrated in Figure 4 (next page). The characteristics in red increase the value of the subject above the average value and those in blue lower value. In this case, the fact that the subject property is good construction quality (7.5) and large (2,775 square feet) increase value the most and the fact that the subject sold early in the eight year sale period (month 7) decreases value the most. Of course, for value prediction purposes we would either use time-adjusted sale prices or set sale month to the valuation date.

Figure 4. Force Plot

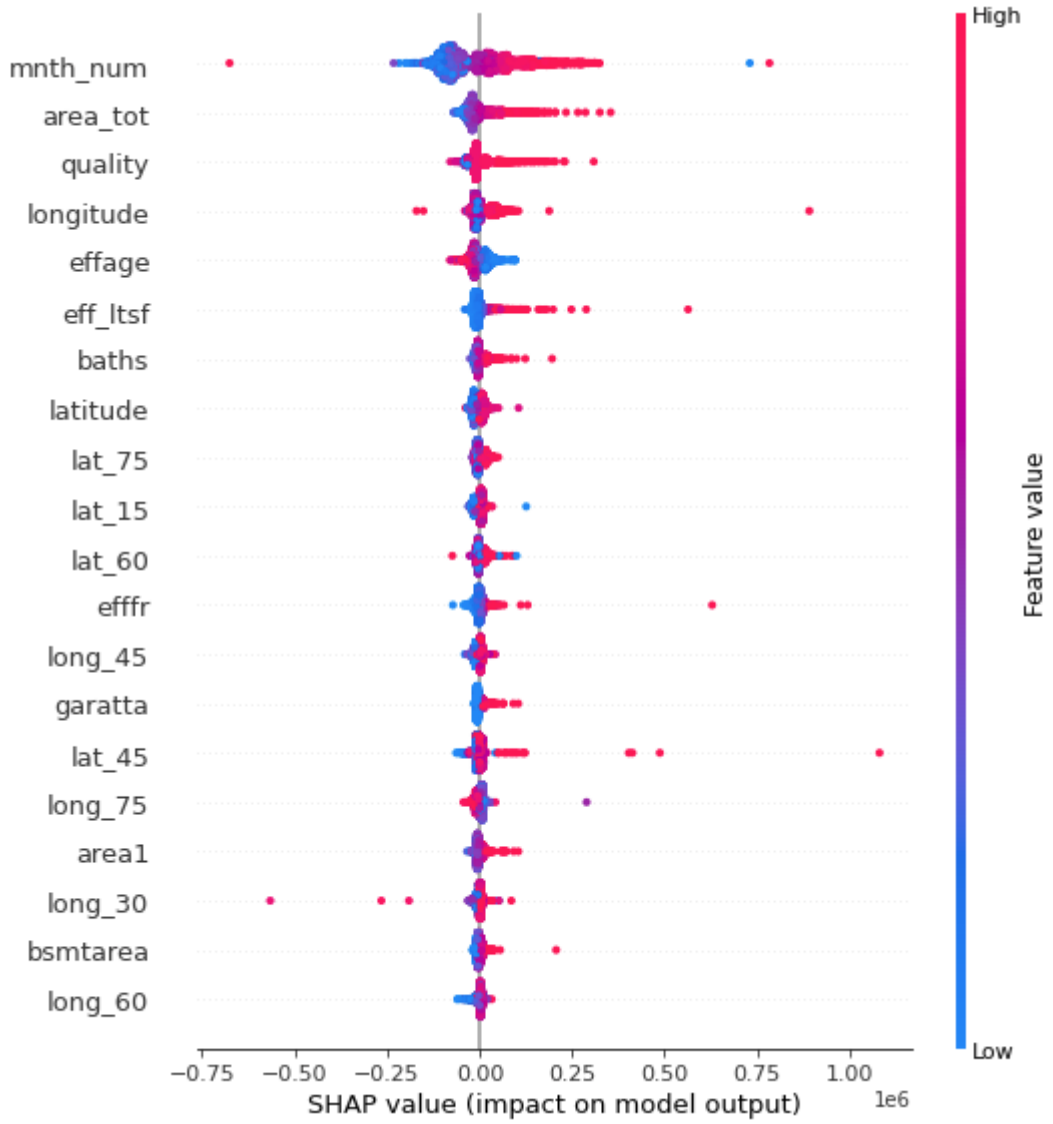
Source: Courtesy of Municipal Property Assessment Corporation (MPAC), Ontario, Canada.

Shapley values thus provide a means of showing the impact of a subject property's features on value without pre-ordering the variables. Of course, the required calculations are computer-intensive. In our illustration, the calculation for garage would have to be repeated many times and we would have to do the same for other variables of interest to obtain the relative influences illustrated in Figure 4. This is unlikely to be an issue for an individual property, but can be time-consuming if applied in batch mode to a large number of properties.

Finally, value influence plots and Shapley values can be combined into "summary plots" as illustrated in Figure 5 (next page) for the 20 most important variables in an AI model. Because values were rising strongly during the sales period, which spanned eight years, the variable for sale date (mnth_num) is most important, followed by living area (area_tot), quality, and so forth.

The dots in the graph are Shapley values for each property with overlaying values jiggered to give a sense of their concentration. Blue dots represent below-average values for a feature and red dots represent above-average values. We can see, for example, that low (blue) values for mnth_num are generally associated with negative impacts while more recent (red) sales yield positive impacts. On the other hand, low values for effage have positive impacts while high values have negative impacts. The graph also reveals some outlier Shapley values, likely due to data anomalies.

Figure 5. Example of Summary Plot



Source: Courtesy of Municipal Property Assessment Corporation (MPAC), Ontario, Canada.

Recall that AI-models are not equation-based but, rather, based on pattern recognition and while, with the exception of LIME methods, the techniques illustrated above do not provide an equation per se, they give a strong sense of what features drive values and what impact they have on predictions, either globally in the case of value influence and partial dependence plots, or on individual values in the case of break-down plots and Shapley values. While these techniques are quite recent and still evolving, if one is comfortable with using AI models as a value-engine, there are effective techniques for illustrating the impact of each relevant property characteristic on generated values.

4. Advantages, Challenges and Technical Issues

4.1 Advantages of AI

AI offers a number of formidable advantages:

- Accuracy. AI is designed to learn from the data (i.e., follow market evidence) so that it can predict probable prices for similar properties. In particular, AI is able to adapt to data nonlinearities and interactions that are difficult to capture in traditional methods. Many tests have been conducted comparing the results of AI methods against traditional appraisal methods, including MRA. Although the results are mixed and dependent on the care and specification of the models against which they are compared, most researchers conclude that AI models can outperform equation-based methods. Of course, what is relevant to an assessment jurisdiction is whether AI will perform better than its current methods and whether that improvement is worth the transition issues. Hence the need for thorough testing on a jurisdiction's own data and consideration of the various advantages, limitations, and issues involved as discussed below.

- Objectivity. AI is totally objective without preconceptions or preferences. While it requires due care in parameter specification, AI obviates logic, transformation, and specification ("human") errors that can confound traditional models.

- Speed. Assuming reasonable parameters have been provided, AI will develop predictions in minimal time compared to that required for an experienced modeler to specify and calibrate an effective equation-based model. Because AI methods are based on splitting the data into more homogeneous branches, it is also possible that AI can reduce the number of models required for accuracy.

- Reasonable cost. As with traditional appraisal methods, AI requires skilled professionals, enabling software, and good data. However, by reducing the amount of time spent in model development, once experience is gained, AI should be cheaper, or at least no more expensive, than traditional methods.

4.2 Limitations and Challenges Associated with AI

There are also significant challenges associated with AI-based appraisals:

- Explainability. While AI is programmed to achieve high accuracy and techniques are available to show the relative importance of variables, there is no tangible equation or coefficients to show how individual values are calculated. As such, AI does not lend itself to simple explanation. Of course, the assessor can employ comparables

in evaluating individual values and supporting them as appropriate. However, both staff and the public are also accustomed to concrete algorithms that can be referenced, reviewed, and critiqued if desired. Assessors will need to consider how AI is to be presented and explained to stakeholders. Value influence plots, partial dependency plots, break-down plots, Shapley values, and other methods that analyse how predictions change as variable inputs are changed can help in this regard.

- Documentation. Related to the above, valuation procedures and models should be documented. Because AI does not produce an equation, rates, or adjustments, documentation will necessarily be more conceptual or process oriented. It should explain the general process, number and distribution of sales used, performance (e.g., sales ratio) metrics, variables considered, and their contribution or importance in value estimation.

- Required skills. AI does not provide out-of-the box push-button appraisals. It requires data scientists trained in the software and skilled in the method(s) being employed.

- Limited appraiser interaction and direction. In applying traditional appraisal methods, a modeler proceeds in a logical sequence, examining and cleaning the data, determining variable transformations, building and testing an initial model, and making refinements. Such an approach is not possible with AI. While appraisers still prepare and clean the data and test results, the modeler is less involved in the calibration process and relies on the machine to determine the predictions.

- Maintenance and allocation issues. While AI is adept at predicting total property values, the assessor must consider such issues as the allocation of value between land and buildings, treatment of dual tax classes, and updating of values for physical changes. For example, if a garage is added to a property, must the entire value be recalculated and, if so, a new analysis run? If two adjacent properties add an identical garage, will their values change by the same amount or percentage?

- Potential for over-fitting. AI methods can be prone to over-fitting, although there are a number of techniques available to minimize and, arguably, eliminate the problem. Provision for a holdout ("test") group or cross validation can be effective in gauging likely performance for unsold properties.

- Change management. Clearly, adoption of AI methods will require a cultural change in the assessment office. Assessors will have to buy into the method and know how to interact with - and explain/defend - the resulting values. They will have to understand how AI uses and analyses data, what data items drive values, and how to identify and address deficient or out-of-line values. Changes to customer

relations materials, appraisal manuals, and training courses should be considered. Staff reassignments and modifications in job responsibilities may be involved.

4.3 Technical Issues

Below are some technical issues that users of AI will want to consider.

- Data preparation. Like all valuation methods, success is dependent on the data used in analysis. Due diligence in sales screening and data quality control are essential. In addition, ensuring that data are in a format amenable for analysis will enhance predictability. For example, it may be helpful to consider eliminating redundant or unimportant characteristics and combining neighbourhoods or other features with too few sales.

- Sample Size and Representativeness. AI requires large samples dependent on the number of independent variables used in the analysis. There must also be enough sales to support a holdout group or cross-validation. While sample sufficiency and representativeness are issues in MRA, they can be even more so in AI because of its use of shallow trees and repeated iterations/refinements.

- Time of sale. Users must ensure that values are predicted as of the valuation date. This can, of course, be accomplished by using time-adjusted prices in the analysis. If unadjusted sales are used in conjunction with time of sale variables, AI may simply produce different values dependent on sale date, which may be acceptable if, for valuation, sale date can be set to the target date. In any case, time trends will be implicit rather than apparent, so that transparency would be improved by determining a way of illustrating, and possibly quantifying, time trends. Partial dependence plots may help visualize the effects of time on value.

- Parameter dependency. Just as other valuation methods are dependent on modelling decisions made by the appraiser, the success of AI models will vary significantly with selected methods and parameters. Important parameters can include the number of trees (iterations), percentage of observation used in each tree, tree depth, learning rate, and validation method.

- Appraiser interaction and quality control. Because there is no equation or tangible rates and adjustments, appraisers will want to use alternative means of ensuring that values are reasonable, equitable, and accurate. As always, primary tools in this regard are the ratio study, value change analysis, and field and/or desk reviews. But what if results indicate issues, such as newer residential properties being under-valued or particular types of property, e.g., those adjacent to municipal golf courses, being over-valued? In MRA, the modeler can devise appropriate transformations to address such issues, override selected coefficients, or apply

outside-the-model adjustment factors. Such options appear more limited with AI, although adjustment factors could be considered. In any case, it is important that appraisers test predicted values thoroughly and ensure their reasonableness and uniformity.

- Confidence measures. In addition to confidence measures related to overall values, additive and multiplicative MRA models will produce confidence intervals about individual value estimates. While, as a general matter, it is well known that prediction accuracy varies between typical and atypical properties, users of AI may want to consider what overall accuracy measures are available and whether the techniques being considered will generate reliability measures for individual estimates. In addition to parametric measures, consideration can be given to such factors as the density of sales in the vicinity of the subject, the similarity of the subject to those sales, and the variance of prices in the area.

- Reproducibility. Desirably, values should be reproducible so that re-running the procedure on the same dataset produces the same results. Users may want to determine whether their chosen method and software has this ability, e.g., by saving the model, setting a "seed" value, or other means.

- Deployment. Like most MRA applications, AI resides in an external software package so that it must be integrated with the CAMA system for valuation. While values can be simply exported to CAMA, users will want a way to update values in real time when data changes are made to properties.

5. Conclusions and Guidance

5.1 IPTI Conclusions

Based on its research, IPTI draws the following conclusions:

- With the appropriate expertise and deployment strategy, AI may produce values that are more accurate than those obtained by traditional equation-based approaches. Additional advantages include objectivity, efficiency, and cost.
- AI methods are widely used in private sector appraisals and are likely to get increased consideration, testing, and potential use in the public sector, either as a primary or supplemental valuation method.
- Issues associated with AI include the lack of a tangible equation, user understanding and acceptance, and CAMA system interface.
- Because of AI's susceptibility to over-fitting, use of a holdout group or cross-validation is essential. Confidence measures, which may be statistical or heuristic based, can help evaluate the reliability of individual predictions.
- As in other appraisal methods, success depends on data sufficiency and integrity. In addition, a jurisdiction must possess the necessary data analyst(s) and supporting software.
- AI-based values need not be standalone values but can be compared against, and used as a check on, more traditional values. Once an assessor has gained confidence with AI and addressed relevant issues, it can be given more prominence or used as the primary valuation method.
- AI may not be for everyone. While some jurisdictions will find that the benefits of AI offset its drawbacks, others are producing accurate and equitable values using current methods, lack the required technical expertise, or simply prefer their current system. In short, AI provides a powerful option for those who have the disposition to use it and the data scientists to test and potentially adopt it.
- IPTI recommends that the assessment industry consider and debate the guidance offered in this paper as part of a process to have AI adopted on a more formal basis, either as a standalone standard or as an addition to an existing standard. Assessors can learn much from AI applications for private sector appraisals, as well as from the experience of jurisdictions pioneering and testing the use of AI methods. Particularly helpful, however, would be a peer-approved standard or guide on the development and deployment of a credible, properly tested, and defensible AI-based appraisal system.
- The development of such a standard is a key objective of this white paper and IPTI recommends that the assessment profession define a concrete strategy for producing an AI standard that will guide assessors who are interested in exploring - and potentially adopting - this promising technology.

5.2 IPTI Guidance

IPTI concludes this white paper with a brief summary of what it considers to be guidance on the essential requirements and components for a successful AI implementation:

- A jurisdiction must have the necessary management commitment and staff buy-in. Changes in the organization's structure and job responsibilities may be required.
- A jurisdiction must have, or contract for, the necessary technical expertise and enabling software.
- Data integrity and preparation are paramount.
- Determination of optimal methods and parameters requires testing and analysis.
- Steps should be taken to avoid over-fitting; use of a holdout group or cross-validation is essential.
- Attention should be given to such technical issues as outliers, time-of-sale, seldom-occurring features, and atypical properties.
- Predicted values should be scrutinized through traditional ratio studies and procedures should be adopted for handling problems or inequities that are found.
- Procedures should be in place for an explanation of the chosen valuation approach, demonstrating equity, and evaluating and supporting values through comparable sales and/or other accepted procedures.
- Potential users should explore and learn from the experience of other jurisdictions who are using, or have tested the use of, AI.
- Jurisdictions need not take a "leapfrog" approach. They can use AI as a supplementary method and optionally increase reliance on it once they have gained experience and confidence in it and addressed related appraisal and change management issues.

Selected References

Alwosheel, A., van Cranenburgh, S., & Chorus, C. G. - "Is Your Dataset Big Enough? Sample Size Requirements When Using Artificial Neural Networks for Discrete Choice Analysis". Journal of Choice Modelling, 28 (2018).

AVM News. Published quarterly by Automated Value Metrics. www.avmetrics.net

Paul Bidanset - "Moving Automated Valuation Models Out of the Box: A Global Geography of AVMs". *Fair & Equitable* (July 2014).

Biecek, Premyslaw and Burzykowski, Thomas *Explanatory Model Analysis*. Chapman and Hall/CRC (New Yok, 2021). https://pbiecek.github.io/ema/

Borst, R.A. - (1995) "Artificial neural networks in mass appraisal", *Journal of Property Tax Assessment & Administration*, Vol. 1 No. 2, pp. 5-15.

Jason Brownlee - "Classification and Regression Trees for Machine Learning". Understanding Machine Learning Algorithms (April 2016).

Buodd, Martin and Deras, Erland 2020 – "Machine Learning for Property Valuation: An Empirical Guide to How Property Price Predictions Can Improve Property Tax Estimations in Norway". Master's Thesis. Norwegian School of Economics.

California, University of - UC Business Analytics R Programming Guide (June 2018)

N. B. Caphalkar & Sayali Sandbhor - "Use of Artificial Intelligence in Real Property Valuation." International Journal of Engineering and Technology (June-July 2013).

Do, A.Q. and Grudnitski, G. - (1992) "A Neural Network Approach to Residential Property Appraisal", *The Real Estate Appraiser*, December, pp. 38-45.

Niklas Donges - "The Random Forest Algorithm." Towards Data Science (Feb 2018).

Mary Lou Downie & Gill Robston - Automated Valuation Models: An International Perspective. Council of Mortgage Lenders (October 2007).

Dzurllkanian, Z.D., Kamarudin, N. and McCluskey, W.J. - (2014) Boosted regression trees: an application for the mass appraisal of residential property in Malaysia, *Journal of Financial Management of Property and Construction*, 19(2), pp. 152-167.

Elith, J., Leathwick, J.R. and Hastie, T. - (2008) "A working guide to boosted regression trees", *Journal of Animal Ecology*, Vol. 77, pp. 802–813.

Julie Elston & David Purdon - "An Expert Systems Interface". SAS Users Group International (Proceeding of the 1988 Conference).

Guru99 2021 – "Back Propagation Neural Network: What is Backpropagation Algorithm in Machine Learning"; www.guru99.com

Lars Hulstaert - "Understanding Model Predictions with LIME", Toward Data Science (July 11, 2018).

International Association of Assessing Officers - Standard on Automated Valuation Models (2018).

International Association of Assessing Officers - Standard on Mass Appraisal of Real Property (2013).

International Property Tax Institute - "Review and Recommendations for Implementation of Artificial Intelligence Models". Report Prepared for Property Valuation Services Corporation, Province of Nova Scotia (April 2019).

Hans Isakson, Mark Ecker, and Lee Kennedy - "Principles for Calculating AVM Performance Metrics". *The Appraisal Journal* (Winter 2020).

David Jensen - "Artificial Intelligence in Computer-Assisted Mass Appraisal". Property Tax Journal (March 1990).

Lee Kennedy, Mark Ecker, and Hans Isaakson. "AVM Testing and Evaluation using Performance Metrics". May 2021. https//orcid.org/0000-0002-6914-2432.

Patrick Kiger - "The Developing Power of Big Data". Urban Land (Fall 2018).

Nils Kok, Eija-Leena Koponen, Carmen Adriana Martinez-Barbosa - "Big Data in Real Estate? From Manual Appraisal to Automated Valuation". Journal of Portfolio Management (Special Real Estate Issue 2017).

Lenk, M., Worzala, E., and Silva, A - (1997) "High-tech valuation: should artificial neural networks bypass the human valuer?", *Journal of Property Valuation and Investment,* Vol. 15, No. 1, pp. 8-26.

Limsonbunchai, V., Gan, C. and Lee, M. - (2004) "House price prediction: hedonic price model vs. artificial neural network", *American Journal of Applied Sciences,* Vol. 1 No. 3, pp. 193-201.

Liu, J.-G., Zhang, X.-L. and Wu, W.-P. - (2006) "Application of Fuzzy Neural Network for Real Estate Prediction", *Advances in Neural Networks*, pp. 1187-91.

Mark Linne - "Artificial Intelligence and the Future of Valuation". Fair+Equitable (January 2019).

Lundberg, Scott and Su-In Lee – "A Unified Approach to Interpreting Model Predictions". Advances in Neural Information Processing Systems 30 (Montreal: Curran Associates, 2017).

McCluskey, W.J - (1996) "Predictive Accuracy of Machine Learning Models for Mass Appraisal of Residential Property", *New Zealand Valuer's Journal,* July, pp. 41-47.

McCluskey, W.J. and Anand, S. - (1999) "The Application of Intelligent Hybrid Techniques for the Mass Appraisal of Residential Properties", *Journal of Property Investment & Finance*, Vol. 17 No. 3 pp. 218-38.

McCluskey, W.J. and Borst, R.A. - (1997) "An Evaluation of MRA, Comparable Sale Analysis, and ANNs for the Mass Appraisal of Residential Properties in Northern Ireland", *Assessment Journal*, Vol. 4 No. 1, pp. 47-55.

McCluskey, W.J., McCord, M., Davis, P.T., McIlhatton, D. and Haran, M. - (2012) The Potential of Artificial Neural Networks in Mass Appraisal: The Case Revisited, *Journal of Financial Management of Property and Construction*, Vol. 17, No. 3, pp. 274-292.

McCluskey, W., Dyson, K., McFall, D. and Anand, S. - (1996) "Mass Appraisal for Property Taxation: An Artificial Intelligence Approach", *Land Economics Review,* Vol. 2, No. 1, pp. 25-32.

McGreal, S., Adair, A., McBurney, D. and Patterson, D. - (1998) "Neural networks: the prediction of residential values", *Journal of Property Valuation and Investment*, Vol. 16 No. 1, pp. 57 – 70.

Molnar, Christoph – "Interpretable Machine Learning: A Guide to Making Black Box Models Explainable" (2021). https://christophm.github.io/interpretable-ml-book/

Molnar, Christoph – "Shapley Values". https://christophm.github.io/interpretable-ml-book/shapley.html

Pierlugi Morano, Francesco Tajani, Maria Torre Carmelo - "Artificial Intelligence in Property Valuation." Advances in Environmental Science and Energy Planning (2015).

Nguyen, Darla 2020 – "Explain Your ML Model Predictions with Local Interpretable Model Agnostic Explanations (LIME)". https://medium.com/xebia-france/explain-your-ml-model-predictions-with-local-interpretable-model-agnostic-explanations-lime-82343c5689db.

Nguyen, N. and Cripps, A. - (2001) "Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks", *Journal of Real Estate Research*, Vol. 22 No. 3, pp. 313-336.

Ray Palmer, Eric Yu Wang, Clyde Dent - "An Expert Systems Application for Statistical Techniques". SAS Users Group International (Proceeding of the 1989 Conference).

Julia Nunez-Tabales, Francisco Rey-Carmona, Jose Caridad y Ocerin - "Commercial Properties Prices Appraisal: Alternative Approach Based on Neural Networks." International Journal of Artificial Intelligence (vol 14, 2016).

OECD - Principles on Artificial Intelligence. OECD (2019).

Peterson, S, and Flanagan, A.B. - (2009) "Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal", *Journal of Real Estate Research*, Vol. 31 No. 2, pp. 147-164.

RICS - "The Future of Valuations: The Relevance of Real Estate Valuations for Institutional Investors and Banks - Views from a European Expert Groups". RICS (Nov 2017).

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin - "Model-agnostic Interpretability of Machine Learning" International Council for Machine Learning Workshop on Human Interpretability in Machine Learning (2016).

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin – "Why Should I Trust You? Explaining the Predictions of Any Classifier". *Proceedings of the 22nd ACMSIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Memory, 2016.

Peter Rossini - "Using Expert Systems and Artificial Intelligence for Real Estate Forecasting". Sixth Annual Pacific-Rim Real Estate Conference (Jan 2000).

Harshdeep Singh - "Understanding Gradient Boosting Machines". Towards Data Science (Nov 2018).

Tay, D.P.H. and Ho, D.K.H. - (1992) "Artificial Intelligence and the Mass Appraisal of Residential Apartments", *Journal of Property Valuation and Investment*, Vol. 10, No. 2, pp. 525 – 540.

Wall Street Journal - "The Future of Housing Rises in Phoenix". WSJ (19 June 2019).

Worzala, E., Lenk, M. and Silva, A. - (1995) "An Exploration of Neural Networks and its Application to Real Estate Valuation", *Journal of Real Estate Research*, Vol. 10 No. 2, pp. 185-201.

Josep Zuranda, Alan Levitan, Jian Guan - "A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Contest". Journal of Real Estate Research, vol 33, no 3 (2011).