A Review of the Methods, Applications, and Challenges of Adopting Artificial Intelligence in the Property Assessment Office

(2022). A Review of the Methods, Applications, and Challenges of Adopting Artificial Intelligence in the Property Assessment Office. *Journal of Property Tax Assessment and Administration*, *19*(1), 21-80.

Published in:
Journal of Property Tax Assessment and Administration

Publication Status:
Published (in print/issue): 20/06/2022

Document Version
Publisher's PDF, also known as Version of record

A Review of the Methods, Applications, and Challenges of Adopting Artificial Intelligence in the Property Assessment Office

BY THE ARTIFICIAL INTELLIGENCE TASK FORCE

Summary

In early 2020, IAAO President Amy Rasmussen created the Artificial Intelligence Task Force with the goal of developing a white paper describing the impact and uses of artificial intelligence (AI) in government valuation offices.

The COVID-19 pandemic in early 2020 forced government valuation offices to adapt overnight. Many jurisdictions rapidly virtualized tasks and duties, which accelerated ongoing efforts to utilize office automation and implement intelligent software solutions. More and more, workflows incorporating digital information and multiple sources of data are processed and analyzed using software and integrated applications. The fully integrated workflows facilitate the increased usage of AI in operations, assessment, and valuation.

This white paper delivers an introduction and overview of AI through case and pilot studies and review of relevant analytic methods while touching on possible organizational impacts. The paper looks at the changing role of valuers and assessment administrators and the evolution of valuation offices where AI will be used to improve operations, value estimates, and administration. It provides illustrative examples of AI use in the conduct of tax assessment, including the administrative aspects not directly involved in valuation. While there is substantial fanfare around valuation with AI, many of the benefits to be

Acknowledments

The Artificial Intelligence Task Force developed the paper and chairs during development were Margie Cusack, FIAAO, and Carmela Quintos, Ph.D. Members who served on the Task Force were Kristie Foster; Nik Lomax, PhD.; Michael McCord, PhD.; Ashley Wu, Ph.D.; Cory Yemen, Ph.D.; Joseph Wehrli; and Carl Zimmerman, Ph.D. Also assisting were Paul Bidanset, Travis Horne, and special mention to William McCluskey, Ph.D. of the University of Ulster.

realized from the technology are in areas of administration, validation, and oversight. This is reflected in the case studies included, with more than half involving AI applications outside of the explicit valuation function.

The introduction provides a definition and brief history of AI. It also helps disentangle the raft of AI methods with how they are used and provides a concrete list of which assessment activities may benefit from those general classes of algorithms. More importantly, the first section helps put into context why AI is becoming more widespread and what that means for organizations from both staffing and administration standpoints.

After the introduction's overview of what AI is, why it has captured professional imagination, and the organizational changes it portends, we provide examples of current uses by assessing organizations and their partners. The first case study is about the Property Valuation Services Corporation's (PVSC) foray as the first organization in Canada to publish a tax assessment roll using AI-based valuations. This case study highlights the multiyear process leading the organization to that accomplishment and the lessons it learned along the way.

The second use case is a pilot study by PVSC. The section discusses the success of AI, particularly machine-learning methods, for the valuation of residential properties in the Netherlands. The third use case is from BC Assessment (BCA) and describes how valuations of manufactured homes were conducted using AI methods. For successful adoption of AI in an assessment office, this case study highlights the importance of communication and feedback from appraisers and integration of AI-modeled values with the computer-assisted mass appraisal (CAMA) system.

The fourth case study comes from the City of New York and showcases applications using AI to better manage form intake and processing. Using optical character recognition (OCR), it is possible to process the volumes of senior exemption applications and condominium declaration forms received in paper and PDF formats. As with the other case examples, the results still require human oversight but provide a significant improvement over the existing process.

The fifth use case also comes from the City of New York. This section discusses how geospatial data and AI methods are being integrated and leveraged to determine land use, detect building changes, and extract parcel data from images and may be used to automate data collection. This section also gives background on the geospatial data required to leverage land use and building change detection applications, which are growing increasingly familiar and important to tax assessment organizations.

The sixth application involves integrating AI-powered valuation as a feature within CAMA. To illustrate the potential of AI to automate sales-based valuation models, this study examines Tyler Technologies' experience trying to provide an AI-powered valuation option for its users. It also clarifies the technology's perceived limitations, which create headwinds for widespread adoption.

This section ends with a discussion of international adoption of AI in property assessment offices in four African nations: Rwanda, Nigeria, Uganda, and Zambia. The full digitalization of their records and workflows using imagery and modern technology allows them to modernize their systems without going through and updating legacy records and operational processes found in more established assessment jurisdictions.

Following the case studies, the reader will find a section delving deeper into the core machine learning (ML) and AI methods underpinning these applications. ML is covered in the first part of this section. Other methods discussed cover key concepts in artificial neural networks and search and optimization, which underpin virtually every AI application.

Finally, the paper closes with recommendations. Key takeaways are that some tax assessment organizations and their partners are already cautiously adopting AI. The technology's adoption will grow more widespread and touch every tax assessment organization. As such, familiarity with how it is being used, a basic understanding of what is driving these changes, and what they mean for your organization is important.

1. INTRODUCTION

The COVID-19 pandemic has been a transformative event that forced many segments of the economy into a virtual and digital environment for the first time. This disruption presents opportunities to accelerate workplace efficiencies using digital technologies.

The combination of the explosive growth of information, digitization, nearly unlimited computer power, and the use of artificial intelligence (AI) is unleashing new ways of working with and analyzing information and data. Valuation and assessment organizations that use digital tools integrated with AI and machine learning (ML) will experience operational impacts as this technology becomes further available and entrenched in public administration systems because of the flexibility and simplicity of these models

Property assessment organizations have deep historical roots when manual recordkeeping, hand-drawn maps, and blueprint-style construction drawings predominated in the real estate and construction industries. The work was centered on now-dated paper and analog methods in construction and real estate not typically associated with the cutting edge of technology. Yet property assessment organizations have all the hallmarks of post-digital, data-driven enterprises as they are becoming dependent on diverse types of digital data (e.g. geographic data and database-centered financial records), data modeling and analysis, computing software and hardware, 3D digital visualization, and data science methods involving AI and ML.

Property assessment organizations are evolving and managing a confluence of technological, scientific, and organizational changes that have emerged in the past five to 10 years during the transition from nondigital to digital workflows. The new processes wrought by the digitization of data and resulting organizational changes require new skill sets, tools, and theoretical approaches, but they can potentially bring large benefits in the efficiency of the assessment organization. AI adoption and uptake for operations, assessment, and valuation by public administration organizations is probably inevitable in some form because of the analytic power of these tools. AI models will require predictability and reliability in their outputs, some sort of transparency for the analytic and valuation processes, an internal and external understanding of AI methodologies, and institutional values like fairness, equity, and public trust so that these tools and methods are not used incorrectly or abusively.

This white paper is intended to give the reader a background in the concepts and range of applications of AI in the assessment office. AI is adaptable and has the potential to increasing efficiencies and accuracy. The objectives of this white paper are to:

1. Introduce readers to AI and explain the digital ecosystem that is fostering the rapid development of this technology
2. Familiarize readers with a range of relevant assessment and valuation case and pilot studies to facilitate further projects among practitioners
3. Introduce the most important technical methodologies, terminologies, and

concepts in AI that apply to valuation, such as ML, neural networks and deep learning, and search and optimization

4. Introduce a framework for how AI will be incorporated into CAMA software
5. Guide the reader forward with recommendations for a path to AI evaluation and adoption.

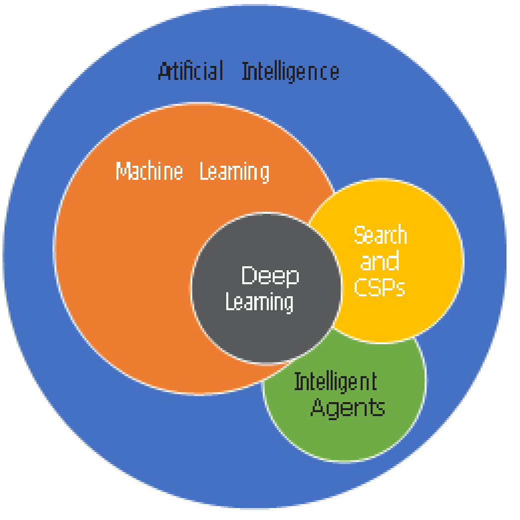
1.1 AI Definition and History

AI is a term with both colloquial and profession-specific meanings and relates to human-like problem-solving, intuitive ML algorithms, and optimization methods that are particularly well-suited for finding patterns in data. It is an amorphous term even for professionals in the field. The term originated in the 1950s and was used to describe a range of intelligence and learning possibilities in machines, including approximating human language; storing and utilizing input they receive; asking questions and creating insights; showing adaptability and pattern identification in new contexts; recognizing speech; and displaying the ability to manipulate objects (Russell and Norvig, 2020). Even among today's data science and AI professionals, a precise and meaningful definition is still elusive; some are mathematicaland computer science-oriented, while others focus on intelligence and cognition issues.

The popular press and imagination typically associate AI with convenience features in consumer technology enabled by a narrow subset called "deep learning". These are not dissimilar to Russell and Norvig's attributes of AI. Whether it is Siri's voice recognition on an iPhone, Tesla's self-driving car, Google's translation capabilities, or Facebook's ability to suggest which friends are in an image, features like these are relatable to us. AI is being embedded into more mundane operational activities in businesses. For instance, an airline's scheduling application or a package delivery route optimization system can be powered by AI. One of the case studies uses AI to extract information from PDFs.

Note to the reader: When we use the term AI, we are condensing several related areas because AI is a broad term that, in addition to includes subfields such as machine learning (ML), deep learning (DL), data mining, and search and optimization, that we will touch on in this white paper (Figure 1). ML, for instance, is defined as statistical methods or algorithms, typically implemented in computers, that can find patterns in massive amounts of data. These tasks include classification, automated pattern recognition, search (found in the computer science literature), and prediction methods. It encompasses a wide and varied list of methods, including neural networks. Unfortunately, the scientific and professional literature has many different definitions. When we use the term AI in the rest of this paper, we are referring to the full range of ML and AI techniques applicable to tax assessment problems.

Figure 1. Taxonomy of AI with Subfields



A useful definition of AI for this white paper includes a combination of ML and AI applications, concepts, and algorithms. We focus this definition on the components most relevant to the assessment and valuation domain. Our definition: AI is the collection of mathematical and statistical methods operationalized in software allowing for outputs to be generated absent formal programming from inputted data.

AI is a collection of methods that enable a system to provide an optimized response or solution to a problem it has not been specifically instructed how to deal with. The model is created by learning relationships from the specified training data. More vividly, it is being able to describe depreciation without requiring a formal schedule or mathematical model; it is being able to translate English to French without linguistic rules; it is being able to automatically identify the building material called "brick" in a new photograph taken by an assessor; and it is automatically extracting text from a handwritten document.

Whatever the definition, organizations, including those in public administration, are using these tools for critical human systems and decision-making that impact our lives in profound ways, such as prison sentencing, fire prediction, allocation of unemployment benefits, driverless cars, financial services, and, of course, tax assessment, modeling, and valuation.

Though AI tools are powerful, transparency and interpretability are difficult because the AI model is a black-box system where the relationships and functions between the input variables and outputs are unknown, and it lacks the symbolic structure found in statistics that people are used to. Critically, the underlying data used to train the model becomes important because biases in the data can end up in the model. Organizations will need to think carefully about equity and fairness issues with these models and about how to present them to their citizens and taxpayers. As we will see later, the assessment software can be organized and designed to help improve transparency and usability for the end user.

Many familiar problems and operational issues for assessment organizations can be potentially addressed with AI. Areas that could benefit include:

- Data collection of physicals
- Error corrections and process automation
- Community, land, and tax administration
- Land and economic development
- Valuation and assessment
- Billing, collection, and revenue

Within assessment and valuation, specific areas that are ripe for AI implementation and application include market modeling, comparable analysis, performance and equity analysis, data aggregation, and justification/defense.

1.2 A Taxonomy of AI

In a sense, the higher goal of AI is that a user states the problem to the AI model, or software, or algorithm, and then the "artificial intelligence" determines how to solve the problem using an optimized solution that meets the user's business requirements.

AI is broad, and the taxonomy of AI covers several subfields (figure 1). This document will focus mostly on machine learning and deep learning, which are structurally similar and more familiar to most readers; and search and constraint satisfaction problems (CSPs), which come from computer science theory. Intelligent agents are another subfield but are not covered in this document.

ML is the tent under which a vast set of algorithms exist that allow computers to learn from the data they are exposed to and trained on. These methods may have statistical motivations (e.g. normal distribution), but that is not a requirement. This data can come in the form of text, numbers, images, sound, video, maps, and sensor readings. There are a variety of algorithms, and they generally fall into one of four conceptual application areas:

- Regression, which provides a continuous number as an output (e.g. property value_, given some inputs
- Classification, which is providing a class or group assignment (e.g. building class) when classes or groups are predefined, given some inputs
- Clustering, which creates related groups and assigns the input to one of those groups when they are not initially defined. Generation, which is less well known and creates stylized outputs like images or sounds based on inputs like old maps or an artist's personal style.

Classical ML methods include decision trees, k-nearest neighbors (KNN), principal components analysis (PCA), support vector machines (SVM), random forests, gradient boosting machines (GBM), and artificial neural networks (ANN), to name a few. Text analytics and natural language processing (NLP) algorithms also fall under the ML tent, as do optical character recognition (OCR) and deep learning recurrent neural networks (RNN), the current champion of speech recognition and translation.

There is substantial overlap in these subfields. CSPs are integral to ML, DL, and key aspects of intelligent agents. If the AI algorithm is based on data, then search and CSPs are used in some form to arrive at the model weights. An overview of several search and CSP methods are provided later.

Deep learning is a specific subset of ANN that has grown rapidly in importance in the last 15 years. It refers to a subfield that uses very large ANNs, solved by search and CSPs, to provide regression, classification, clustering, and generative outputs. DL revolutionized the ability of computers to take raw images or videos and create meaning from them. For instance, it has been used to determine camera depth and car distance, a key property in powering self-driving vehicles, and is critical for recent speech and text recognition improvements. In the assessment world, even multiple regression analysis (MRA) uses a very simple search and CSP algorithm to determine the weights. As the search/CSP and ML methods improve, so too do the capabilities of the intelligent agents built to combine those results.

2. EMERGING TRENDS

Data-driven organizations have used computers to store, analyze, and organize their work for over 50 years, with much of the early work in banking, engineering, multinational corporations, and military organizations. However, starting in the 2000s and accelerating in the past 10 years, a combination of factors — including the rise of the internet and social media, smartphone usage and mobile computing, scalable cloud computing, and data science using AI and ML — intermingled and created a new organizational and analytic model: a new digital ecosystem.

Fully digital companies such as Google, Facebook, Zillow, and Amazon were the first to face business problems created by their exploding volumes of data and extensive data processing needs related to the internet. They needed to adapt business models that could handle their tight integration, feedback, and reliance on the internet, where unstructured digital data was displayed, ingested, and analyzed. They began to use AI methods at scale to find patterns in the large, rapidly changing data streams.

The management, analysis, and processing of these huge data streams have created a technological ecosystem that can handle and analyze data in various formats and have facilitated a positive feedback loop between data volume and analysis, hardware, and software. Hardware and software have greatly improved and adapted to meet the challenges of handling massive data sets. The use, analysis, and management of large data sets with AI is propagating across the entire economy. Any industry, including property assessment and other public sector areas, needs to evaluate how to incorporate these technological and analytic trends. This white paper will provide a background to understanding these changes, techniques, and methods and how they may apply to your organization.

2.1 Digital Ecosystem

2.1.1 Digitization

Digitization is defined as the conversion of analog data such as text, pictures, or sound into a digital form that can be processed by a computer. Analog data is represented in a physical format, while digital information is data consisting of symbols in the form of discrete “0” and “1” values. In an assessment office, an architectural drawing on paper or Mylar is in an analog format, while the same drawing can be encoded in software as a computer-aided design (CAD) file — a digital representation of the same document. To change one part of an architectural drawing on paper you would need to redraw the entire sheet, but once it is in a digital format, different types of automation, visualization, and processing flexibility are unleashed. You can update formats and scales, change one object at a time without updating the whole image, and automate information extraction across your entire library of CAD files. The transformation of data from a physical form to a digital form fundamentally changes the data and your workflows because the digital format facilitates data analysis, data integration, speed, and data volumes. Digital data can be processed at scale using computers and software, and can be analyzed using AI and ML.

Assessment and public administration offices receive information in both analog and digital formats. Customers fill out forms on paper (analog) in the office, or they submit information through the website (digital). They might submit parcel information via paper maps from a professional surveyor (analog) but use data from a geographic information system (GIS) to find neighborhood information (digital). They could send photographs of their property from their smartphones (digital) and PDF scans of documents (starts analog then is converted to digital). Assessors also may be working in both analog and digital formats. They might use a database in the office to manage their valuation data and use paper maps in the field to collect information about the properties of interest while reviewing paper records. All digital formats allow for automated processing, workflows, and analyses. Data types can be intermingled to a far greater degree than physical data can be, potentially yielding efficiencies and cost reductions for organizations.

2.1.2 Big Data

The quantity of data in the world is rapidly accelerating. The conversion of analog to digital data, the creation of new digital information from sensors, the growth of smart devices, and the processing of digital data to make new data all contribute to the flood. Communication data, for instance, is proliferating because of social media platforms, smartphones, texting, and emails. On the microblogging site Twitter, many tweets are not done by people but rather by bots, another type of information accelerant. Facebook now has 2 billion users posting, chatting, and sharing videos. Digital cameras on phones continue to improve with ever greater pixel densities, multiple cameras, and data-intensive 4K and 8K video capabilities — all of which produce lots of data

The geospatial and GIS industries are undergoing similar processes with ever larger data sets and more frequent data capture from satellites and drones. It has long used AI and ML. Because geospatial, GIS, and CAD data are embedded with location and geometric information they are good for creating virtual models and visualizations of buildings. The virtual replication of the built environment called digital twinning offers the same advantages of other types of digitization and can replace analog processes for physical inventories and reviews. We include a section on GIS because of the ongoing importance of geospatial data to the assessment office and its early incorporation of AI and ML methods.

2.1.3 Improvements in Computer Hardware and Cloud Computing

The computer hardware ecosystem includes storage, processing, networks, and graphics. These components continue to have exponentially increasing capabilities to process, access, and store ever larger amounts of data. The data has also become less structured and more complex so AI and ML are playing a central role in sorting, classifying, and utilizing this data.

The new computing ecosystem in tax assessment organizations is migrating off the desktop PC or closet server to become more like a utility. Called cloud computing, it is storage and processing available through the internet and sourced to offsite data centers. It represents a paradigm shift because by taking computing off the desktop, it pushes maintenance of the hardware out of the organization. Hardware becomes a service that you can turn on and off, and organizations benefit from the pooling and “virtualization” of computer resources, as opposed to buying increasingly large computing and storage mechanisms. The advantages of cloud computing include limited maintenance and upfront purchasing, flexibility, and, most critically, the ability to rapidly scale processing (including AI processing) depending on the demand to the system. Cloud computing is a critical component in the new AI ecosystems because it facilitates the data-intensive process of training and optimizing data sets with unlimited scaling. The deployment of computing resources for AI is a simple matter of selecting new virtual machines when you reach the limit of the assigned computing configuration.

2.1.4 AI and Data Science

The analysis of data using statistics in industry, academia, and property assessment is nothing new. What changed in the 2000s was the sheer volume of the data and the speed with which it was released due to the build out of the internet and the advent of the smartphone. Companies such as Google, Facebook, and Twitter had billions of images, texts, and files being produced or uploaded within their organizations, and much of data was unstructured. These companies already had skills in computation but needed ways to analyze and classify these diverse data sets for consumer uses. Statistical models like regression have been used for years but often required data to fit certain assumptions, such as normality or linearity. They also were not well-suited for data sets with hundreds or thousands of variables that people cannot fully comprehend.

The power of AI comes from its ability to find nonlinear patterns in huge data sets that cannot be easily found with other methods or described by people. For instance, AI has been extensively and successfully used in voice recognition software, which improved substantially when it was switched from an extensive list of rules to an AI approach, and in image classification tasks such as identifying cats in uploaded pictures from social media sites. Analysts are not required to create rules; rather, they set up the structure and let AI find the optimized relationships between data. Ultimately, the use and refinement of AI methods was facilitated because these companies had scalable cloud computing, software skills, and the ability to integrate AI directly into business workflows.

2.2 Digitalization and Organizational Transformation

There is a distinction between digitalization and digitization. Digitization, as we discussed earlier, is the transformation of analog media and information into digital information (structured with 0's and 1's) and is the precursor of the organizational process of digitalization. Digital information facilitates the flow and intermingling of data, while digitalization refers to the use of digital technologies, analytic skills, and workflows that transform a business model or models of an organization. This process started with fully digital companies such as software firms but is now working its way through the economy into more traditional "brick and mortar" businesses such as real estate, construction, and public administration; insurance companies, for instance, now rely on AI-based tools to make decisions about customers and their policies. And as you will see below, it is filtering into assessment organizations.

As digitalization occurs, more parts of the organization may become intertwined with AI for decision-making, analysis, and operations. Different employment skills become important, and new job titles may mix data science, geospatial analytics, and assessment. Typical tasks that might benefit from AI enablement include process automation and project administration to identify new patterns in large data sets and such as identify fraud or encoding errors for building attribute data. AI-enabled workflows may also allow automated customer engagements where the use of bots, intelligent agents, and other natural language tools are applied to client communication.

An evolutionary pathway for digitalization in an organization is that AI starts in the optimization of big data problems such as a pilot valuation project. AI can then be applied to other operational projects like human resource administration, customer service, and management decision making problems (Fountaine et al., 2019). Organizations that are information-intense and fully digitized, such as financial services, are most involved in AI usage through digital automation. It is easier for them to conduct, test, and pilot projects in the realm of digitalization because their organizations already are awash in this form of data. Organizations can take one of three approaches and should focus on adoption and integration of AI: a hub, a hybrid, and a spoke model. The hub typically has executive-level leadership and is responsible for training and partnerships with data providers and software companies. The spoke involves pushing AI out to the business user for adoption into operations. In the hybrid model, project management, architecture, and IT strategies are often assigned out to the business units. (Fountaine et al., 2019).

2.2.1 Impacts to Organizations and Workforces

Digitization, digitalization, and integration of ML and AI will impact staff and organizational structures. Adapting, training, and educating staff and managers will be important because productivity is increased when the AI and humans work in a collaborative fashion and take advantage of the tools and skills of both. One concept of this collaboration has people performing three roles: train the machines, explain the outcomes, and integrate values. Enhanced and new skill sets will be needed for these roles.

2.2.2 Skill Sets

The use of AI, as it spreads into different aspects of assessment organizations, will require employees to extend their skill sets because output from AI models and applications comes through a black-box function and may not be easy to interpret. The results can be subject to bias based on the types of data the models are trained on, so it will be important to understand the systems that are being modeled and recognize unusual or spurious outputs. Staff will need more skills in open-source automation, data science, database management, statistics, and visualization to understand and manage the system. Becoming familiar with geospatial data and techniques will become more critical as digital twinning and other 3-d visualizations becomes prominent in CAMA systems. Employees will need to understand the underlying theoretical structure of their operations and models to properly apply and evaluate the outputs they receive from their AI-based applications. Consequently, the workforce will also need to learn and respond more quickly, as the AI will update continually in response to new data.

2.2.3 Transparency and Resident Impacts

The clarity, logic, and transparency of valuation and assessment systems for residents become critical issues when AI models are integrated into the data analysis workflow. Assessment organizations need to address their customers' trust issues. External outputs from AI models will become more ubiquitous, and organizations will need to have strategies for managing the expectations and anxieties of the public about their valuations. As is occurring in the social sciences now in the movement called Open Science, end-to-end transparency will become more important. It will not be enough to present the valuation and assessment results; the whole chain of analytic logic, from data to output, will need to be available and verifiable. New best management practices such as allowing models or code to be shared and reviewed will become more common.

2.3 Conclusion

A new ecosystem of digital technologies and hardware has emerged that parallels and interacts with the immense growth in the availability and types of data. Some of this data comes from the digitization of existing data, but some also comes from geospatial data sources, widely available smartphones, the internet, and automated data production. New algorithms and tools using ML and AI can handle the breadth and volume of this data and solve new problems. These tools mimic some types of human cognition and show additional flexibility for new problems beyond traditional statistical tools. To adopt AI, property assessment organizations will need to use the mutual strengths of their workforces and the appropriate AI tools.

3. CASE STUDIES AND PILOT APPLICATIONS

3.1 AI and Mass Appraisal — The Property Valuation Services Corporation Experience

This section contains a series of case and pilot studies that show applications and issues related to the implementation of AI. The first case study describes how Property Valuation Services Corporation (PVSC) in Nova Scotia, Canada, integrated AI into its assessment and valuation process.

In 2019-20, PVSC became the first jurisdiction in Canada to use AI to predict and apply property values on an official assessment roll and to understand the impact of the COVID-19 pandemic on residential and commercial assessments in Nova Scotia.

3.1.1 The Road to Innovation

PVSC is the independent, not-for-profit organization that assesses all 630,000 properties in Nova Scotia and provides an annual assessment roll to each of the province's 49 municipalities.

PVSC is committed to fostering a modern organization with a progressive vision for property assessment and to using contemporary digital assessment tools and imagery. In recent years, the organization took deliberate steps to realize its vision, including transitioning to a primarily work-from-home model, using Pictometry (angled or oblique digital imagery) to complement field inspections, deploying the MobileAssessorSM app to all field assessors, building an application programming interface (API) as a bridge between the ML tools and iasWorld CAMA system, and building a comparable sales app to assist with sales comparison and value defense.

3.1.2 AI and Property Assessment

In 2017, PVSC executives attended a training session led by the University of Toronto's AI team that highlighted how AI can empower data analysis and prediction. It was then that PVSC began to imagine how AI could advance mass appraisal and the work of property assessment. With the endorsement of its board of directors, PVSC's data science team got to work, exploring using ML for assessment and making it a central part of the organization's strategic plan. Within two years, PVSC would go from idea to implementation.

3.1.3 Harnessing Machine Learning for Mass Appraisal

Beginning in 2018, the data science team partnered with the Department of Statistics at Dalhousie University in Halifax, Nova Scotia. They spent a year methodically evaluating property assessment data, building, testing, and successively improving multiple ML algorithms to determine the best fit for PVSC's data and purpose.

Using these algorithms, they were able to analyze massive quantities of data quickly, identifying patterns and making predictions, as a promising next step to improve the efficiency and accuracy of property assessments.

Following industry best practices, experienced assessors carefully reviewed the ML values, evaluated them against IAAO standards through comprehensive ratio studies, and submitted them for audit by external experts. Once PVSC was satisfied that its ML approach could consistently produce accurate property assessment values, it shared the results with its assessment industry peers at IAAO conferences.

3.1.4 Putting Machine Learning into Action

Initially, PVSC worked with BC Assessment (BCA) on a precursor pilot to evaluate whether AI was a viable valuation method for residential condominiums in BCA's large, active real estate market. PVSC successfully created a single ML model that could do the work of the 1,000 models then used by BCA for condominium valuations. Statistical analysis indicated that the final model performed with an average prediction accuracy within 6.2 % of sales prices.

In 2019, PVSC integrated ML for reassessment into its 2020 assessment roll for Nova Scotia for the first time. Parallel assessment rolls, one using a traditional (market-adjusted cost) approach and the other using an ML (sales comparison) approach, were prepared, and scrutinized. Experienced assessors used statistical criteria and their market knowledge and expertise to compare the two reassessments and identify which accounts would be valued using ML.

In the end, approximately 33% of all residential improved accounts and 30 % of all condominium accounts on PVSC's 2020 assessment roll were valued using AI. The 2020 assessment roll was distributed to Nova Scotian municipalities prior to December 31, 2019, and to property owners across the province on January 13, 2020. During the 31-day inquiry and appeal period that followed, PVSC received:

- 4,781 inquiries regarding residential assessments, including only 88 inquiries regarding ML-valued accounts (or 1.8 % of all residential inquiries)
- 4,962 appeals regarding residential assessments, including only 561 appeals regarding ML-valued accounts (or 11.3 % of all residential appeals)

Ken McKinnon is an IAAO-certified Residential Evaluation Specialist (RES) with PVSC. He describes how ML impacted his work:

> *After 12 years as an assessor, I can say that the statistical results alone show the strength of our ML valuation method but seeing first-hand the time and effort that it saves us is astounding. The data analysis efficiency has given me time to focus on more in-depth market research, develop closer relationships with our municipalities, property owners and agents, and devote extra attention to sales and permit reviews. In the future, I anticipate it'll also give me more time to review additional accounts without activity to ensure comparability.*
>
> *Another big win for me is ML's ability to factor in so many characteristics that we just weren't capable of considering under more traditional methods. At this point, given the efficiencies, sound statistical performance and the ability to evaluate a multitude of variables, I can see ML becoming a foundational tool for property assessments.*

To prepare for the 2021 assessment roll, in the spring of 2020 PVSC built ML models for both the residential improved and condominium files. The ML outputs met all industry standards, but on closer review assessors noticed differences between the ML model values and the market-adjusted cost values in the residential improved file.

Up to this point, improvements to the ML models had been incremental; however, with these discovered differences, PVSC decided to use the rest of 2020 to test different model scenarios to determine an optimal model for residential improved properties. It explored scenarios such as:

- Time trended sales vs. time as an attribute
- Single model vs. market area models
- Many data attributes vs. select data attributes
- Many years of sales vs. select number of years of sales.

As PVSC proceeded with model optimization for the residential improved file, the condominium ML values aligned with market trends and traditional valuations. PVSC decided to use the condominium ML values on its 2021 assessment roll and use the residential improved ML values as a quality check for its market-adjust cost values.

By taking the time to further evaluate and refine its residential improved ML model, PVSC will have an optimal model to use for the 2022 assessment roll that will require only minimal and focused refinement in future years. PVSC also found additional uses for AI in 2020 as a tool to monitor and better understand the impacts of COVID-19 on different property sectors.

PVSC plans to apply ML to more accounts, wherever suitable, in the coming years. While the use of ML as a valuation tool may be new, PVSC's commitment to industry standards and to the oversight and discretion of experienced assessors remains unwavering.

3.1.5 Key Improvements from Machine Learning

At this point, PVSC's experience indicates that ML can provide at least six key improvements for property assessment:

1. Accuracy – PVSC's ratio study results indicate that ML can produce assessment values that achieve the same, or better, accuracy than traditional approaches according to IAAO standards.
2. Efficiency – PVSC's ML reassessment required 85 % fewer person-hours, was completed in less than three months, and allowed for the analysis of greater quantities of property and market data.
3. Explainability – PVSC's inquiry and appeal statistics, as well as anecdotal reports from assessors, indicate that questions and valuation challenges from property owners were handled using available sales information.
4. Expertise – Using ML for a portion of the 2020 and 2021 reassessments allowed PVSC assessors time to focus their experience and expertise on market knowledge, file improvement, judgment, and sound decision-marking rather than data analysis.
5. Flexibility – ML provides the flexibility for reassessment to be completed at any point in time.
6. Strategic Advantage – ML is helping PVSC address organizational challenges

related to budget and resource allocation, as well as emerging issues such as a global pandemic.

The organization's expertise in AI has had the unexpected benefit of helping PVSC better serve its stakeholders and enhance its decision-making ability during the COVID-19 pandemic. PVSC is now doing market analysis and forecasting that would not have been possible without this technology. As a result, PVSC is better equipped to make critical decisions and plan for potential market impacts on residential and commercial sectors in advance of preparations for next year's assessment roll. PVSC has demonstrated that AI will not replace the work of diligent, experienced assessors; it will provide them with improved information for enhanced judgment and better decision-making.

3.2 Machine Learning for the Valuation of Residential Improved Properties in the Netherlands

3.2.1 The Pilot

With a national mandate focused on automation and AI, the Netherlands Council for Real Estate Assessment was eager to investigate and evaluate the effectiveness and viability of deploying ML for predicting property assessments. In 2020, with the Council's guidance, a pilot was formed between PVSC, Belastingsamenwerking West-Brabant (BWB), and CAMA provider 4Value to answer the following questions:

- Is ML a viable tool for predicting property assessments?
- Is there a significant difference in results using time-trended sale prices vs. having time as a variable in the model?
- Is there a significant difference in results between market-area models vs. one model for all BWB?
- Is there a significant difference in results using sales up to a January 1 base date vs. including sales six months after the base date (January 1-July 1)?

3.2.2 Modeling Approach

A ML output is reliant on the input data. The first step of the pilot involved a comprehensive data exploration of BWB's sale and property attributes to determine whether the data was suitable for building accurate ML models.

Once the data was confirmed to be complete and representative and data business rules were established, time-trending and market-area analysis was conducted. PVSC used cluster modeling to identify potential market areas and an IAAO time-trend methodology, which uses regression of sales-to-assessment ratios on time, to establish market trends. Both the time-trending and market-area results were validated by BWB valuation experts with local market knowledge.

Models were built by first separating out a training set and test set (holdout sample) for each model. Different parameter settings were tested to determine which provided the best results for the training data set. Before the model was applied to the full population, its performance was measured through a series of statistical tests on both the training and test sets, and a ratio study was conducted to measure the accuracy and uniformity of the ML predicted values against IAAO acceptance criteria and standards. Once the model was applied to the population, a comprehensive variance analysis was conducted to compare the model results to BWB's traditional assessments.

3.2.3 Model Evaluations

Each model was evaluated using a series of statistics that provided an indication of how well the model was performing (predicted value vs. sale price) using the selected model parameters. The statistical evaluations were conducted on both training and test sets. Emphasis was placed on not only the results of the test set, but also the differences between the results of the training and test sets. This approach allowed for an unbiased evaluation of the ML model by measuring its performance on unseen data.

IAAO provides guidance for measuring the quality of valuations and suggests that a comprehensive ratio study provides the mechanism to evaluate the results of the mass appraisal valuation product from a broad range of perspectives, regardless of the method used to derive the values. A ratio study was conducted on each model to ensure results were within IAAO guidelines.

While the ratio study compared the ML predicted values to sale prices at various levels, it was important to understand the accuracy and stability of the model outputs in other ways. To do this, a comprehensive variance analysis was conducted to compare the model predictions to other model predictions and model predictions to BWB's assessments.

3.2.4 The Results

Over four months, PVSC built 10 models to test model viability, accuracy, and stability. The results indicated that ML is a viable tool for predicting accurate and uniform valuations for residential improved properties in the Netherlands. The results of the time-trend tests showed no significant difference between time-trending sales vs. using time as a variable. Similarly, the results of the market-area analysis showed no significant difference between having defined market areas vs. one model for BWB's full population. The use of six months of additional sales had no significance on the results, indicating that the model is stable and can handle a large range of sales data.

ML is a viable tool for the property valuation of residential improved properties in the Netherlands, and the optimal model for BWB is a single model for its full population that uses time as a variable.

Through this pilot, the Council, BWB, and 4Value gained business and operational insights that align with their focus of finding efficiencies through the automation of processes and adoption of new innovations and technologies.

3.2.5 Considerations

As with any strategic change, implementing ML within the Netherlands would require changes to operational and regulatory environments to accommodate the efficiencies and new practices associated with AI. Incorporating a new approach to value within an existing system should be incremental and well planned. To assist with implementation considerations, the International Property Tax Institute (IPTI) developed an implementation framework to help identify risks, considerations, and mitigation tactics for the integration of AI and ML.

3.3 Embracing AI at BC Assessment: Using AI for Manufactured Home Valuation

This short case study describes how AI was used in a pilot study to address a difficult problem: manufactured home (MH) valuation for a small universe of properties. BC Assessment (BCA) is a Crown Corporation in British Columbia, Canada, that has a legislative mandate to create and maintain uniform property assessments throughout the province. BCA values over 2 million properties across the province on an annual assessment cycle, with a total assessed value of over $2 trillion.

3.3.1 Development of Modeling Modelling at BCA

BCA has a mission to create uniform assessments and trusted property information. Integral to this mission are innovative and resilient data science and analytic teams who embrace continuous improvement and work collaboratively to create innovative solutions and provide insights. As part of this philosophy, BCA created a central modeling team in 2019 — first as a pilot project and now as a permanent department – that consists of appraisers, data scientists, and a manager who all work collaboratively on business problems.

One of the initial goals of the modeling team was to improve the efficiency and effectiveness of BCA's market-adjusted cost approach to valuing properties. In 2019, this meant developing an automated algorithm that integrated with and enhanced appraisers' usual methods of establishing market land rates and market-adjusted cost factors for single family residential properties for the 2020 annual assessment roll (completed in December 2019). This was paired with the adoption of an automated model that time-adjusted sale prices.

In 2020, this single-family residential model was continued and enhanced, and a pilot project was undertaken by the modeling team to value strata residential properties using multiple regression analysis (MRA) models in parallel with the preexisting direct comparison approach. In 2021, the single-family residential model was further enhanced.

3.3.2 Modeling Considerations for Manufactured Homes

MH valuation is a challenge because of the limited number of properties in any one geographical location and the large variance in sale price. Two issues for 2021 were whether to include the setting of market-adjusted cost factors for MHs in the existing single-family residential model or to pair with the existing market-adjusted cost approach of valuing MHs. However, with fewer than 20,000 MH sales in the past five years for a portfolio of approximately 80,000 properties, there were simply too few MH sales in most MH market sets to automate the setting of market-adjusted cost factors with confidence. Another option was to create MRA models for the valuation of MHs. However, this would have involved creating many different MRA models across the province for the wide variety of market sets, which would have taken a good deal of time to train and validate. Again, the relatively small number of sales available in a given market set would likely have posed difficulties. Further, there were some variables that the team hoped to include in a valuation model for MHs that do not lend themselves easily to MRA — for example, geographic location.

BCA saw MH valuation as a relatively low-risk opportunity to pilot how well an AI/ML model could perform on a small group of properties. The MH ML modeling was undertaken in the spring/summer of 2021 by the modeling team in conjunction with business stakeholders; this was done so that preliminary results and feedback could be received before the annual assessment roll calibration ramped up in the fall. If the ML results did not meet the high standards of the business stakeholders, then the preexisting market-adjusted cost approach would be used, with minimal impact on the roll calibration schedule.

3.3.3 Gradient Boosting Model for Manufactured Homes

In past experiments with using ML models for mass appraisal, BCA's data scientists found that both GBM and random forest models perform reasonably well for valuation purposes. Combined with the positive results of the work PVSC did with BCA in 2019 using ML models for strata residential valuation, BCA decided to pilot the valuation of MHs using a single GBM model.

The initiation of the MH ML model pilot had two focuses. The first: to prepare and explore MH data to ensure its accuracy and to perform necessary data transformations (e.g. combining multiple types of porches into a single variable). The second: to communicate with business stakeholders to ensure they were informed of the work being undertaken by the modeling team; to identify with whom and on what schedule the modeling team would be collaborating to build, test, validate, and productionize the model; and to elicit preliminary feedback regarding the pilot process.

All data exploration, data transformation, GBM building, model tuning, and model testing were performed with the open-source R programming language, particularly making use of the tidyverse and XGBoost packages. Data was extracted from BCA's CAMA system databases using SQL queries directly within R. The MH ML model was initially built to estimate market values as of the completed 2021 annual assessment roll, which enabled a direct comparison between the ML model results and the appraiser-valued market-adjusted cost approach values. Once tuned, the initial GBM results were examined by the modeling team, which found that the results matched appraisal knowledge and expectations of variable importance and their impact on market value.

3.3.4 Model Communication and Enhancements

The initial MH AI/ML model results were shared with BCA's MH stakeholders — the AI/ML model-derived values were compared with the 2021 appraiser-derived values, with the observation that values were typically similar between methods. As an additional test, both the ML model and appraiser-derived values were compared with calendar year 2021 sales, which had not been included in either value calibration process. Ratio studies were performed on this holdout sample of sales, with the GBM performing as well as the appraiser-derived values. A Microsoft Power BI dashboard was also created and provided to appraisers so they could examine the ML model results in detail by drilling into ratio study, value comparison, and property details. As a companion to the dashboard, a feedback document was provided that asked several questions about aspects of the ML model to identify possible enhancements or issues. For example, questions were asked about the completeness of the data elements in the model and whether any data transformations were not accurate representations of the data elements. The initial feedback from MH stakeholders provided some enhancements, such as including a new variable in the GBM that transformed MH foundation types.

3.3.5 Integrating the ML Model with the CAMA System

The model-derived MH value estimates were used to instead provide an indicated market-adjusted cost factor for the existing market-adjusted cost approach. The market-adjusted cost factor for each market set was calculated to minimize the residual sum of squared errors between the market-adjusted cost values and the ML model-derived values. By using the market-adjusted cost factor to translate the ML model values to the CAMA system's built-in market-adjusted cost approach, a change in the inventory of MHs would enable real-time value updates by using the existing valuation process that appraisers are familiar with.

In addition to enabling real-time results of inventory changes, integrating the ML model with the existing valuation approach and CAMA system in this way also provides a familiarity to MH stakeholders to ease the transition to the "black box" ML model. Using market-adjusted cost multipliers to translate ML-derived values also enables the use of a process to calibrate the factors that was developed in 2019 for single family residential properties and has already been used for two years successfully. This familiarity with the calibration process, in conjunction with the demonstrated comparability in value estimates and acceptable ratio statistics, has provided confidence in using the ML MH model to set the 2022 annual assessment.

3.4 Digitizing PDF Files Using Optical Character Recognition

In this case study, AI tools are applied to a constrained but significant operational problem facing older tax assessment organizations that need to transform their workflows from analog to digital formats. Some agencies have taken the first step by transitioning from paper to electronic copies (like PDFs) that are scanned. But how do you then digitize them to achieve efficient workflows and processing?

Optical character recognition (OCR) goes a step further by taking the scanned document and converting it so the data is intelligible to AI algorithms. It can involve a variety of techniques, not all of which are covered in this document. This additional step that converts images to intelligible data is a leap that involves a series of advanced programmed engines with algorithms derived from techniques in pattern recognition, AI, and computer vision.

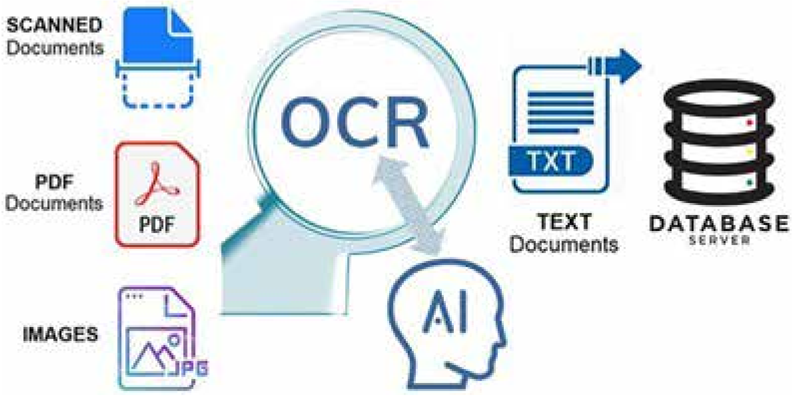
The advantages and efficiencies for digitizing records are clear: the cost of paper storage decreases without the need to manage paper file rooms, communication is more effective and becomes nearly instantaneous as documents are shared, and information security can be improved with digital locks on documents and verification technology. In addition, OCR facilities the most significant changes in business operations where the conversion of paper information enables downstream analytics within scalable databases. Information access and rapid analysis are essential for operational adaptability and success.

3.4.1 The OCR Process

OCR is the process by which scanned documents are analyzed and characters are recognized and made available for editing, searching, and analytics. A basic overview of how an OCR engine processes an image to return text and then saves it into a searchable database is in Figure 2.

- First, the document is scanned by the computer and saved in a PDF or image format
- Second, the image is submitted as input to OCR software. The software matches portions of the image to shapes it is instructed to recognize. Given logic parameters that may include AI techniques depending on the software and application, the OCR engine will make the best guess as to which letter or number a shape represents. At this point, depending on the software or the workflow prescribed by the user, postprocessing checks are done by humans on data the OCR software outputs where the match to known words has a high probability of error. OCR software programs are distinguished by the algorithms used to generate output with minimal error
- Third, OCR results are returned as text in different file formats, including plain text, Word, or Excel
- Finally, the text is saved as a searchable database from which verification and analytics are performed.

Figure 2. The OCR Process



3.4.2 Factors That Affect OCR Accuracy

OCR software can be purchased or is available as open-source software. The degree of accuracy varies significantly depending on the quality of the scanned document, whether it is handwritten or typed, whether it is standardized or free form, and which software is used and its algorithms. The typical OCR process can be split into three fundamental steps: image preprocessing, character recognition, and postprocessing correction.

3.4.2.1 Preprocessing Scanned Images

Because the scanned document is the input to the OCR software, much of the software's performance depends on the scanning process and preparation of the image before the actual character recognition is performed. The scanning device should be maintained regularly to prevent background noise, and all scanned documents should have the same orientation. Most OCR software will correct deficiencies in scanned documents using algorithms based on computer vision techniques. These preprocessing steps aim to correct irregularities from the scanned documents. Techniques include:

- De-skewing: This process detects tilted images and rotates the image so that characters are upright prior to character recognition. Figure 3 is a tilted image from a property exemption application with fictitious information.

Figure 3. Bounding box issue


Section 1: Property Information

Brooklyn 1359 5
123 Atlantic Ave 11H
Brooklyn NY 10011

Type of Property:
Condominium Cooperative 1- to 3-family house 4+ family house or other
If your home has four or more units, enter the % of the space that is used as your primary residence: ____ %
Is any portion of your property used for commercial purposes? Yes No
If yes, enter the percentage used for commercial purposes ____ %
Have you owned this property for at least 12 consecutive months? Yes No
Is the property owned by a trust? Yes No
Did you receive this property through a will? Yes No
Is there a life estate on the property? Yes No
Does a child (including tenants) reside on the property and attend public school in grades pre-K to 12? Yes No

OCR software will identify the bounding box of the image as being tilted and will rotate the image prior to applying character recognition techniques.

- De-noising: This process reduces "noise" coming from scanning devices. Noise may be present as black dots, dark lines, or other artifacts from the scanning device.
- Character Enhancing: De-noising algorithms may lead to some characters losing their edges. OCR software with good performance will enhance character edges, similar to how we enhance or sharpen pictures.
- Histogram Equalization: Parts of an image may appear to be brighter or more exposed. Histogram equalization is a technique that balances out the brightest and darkest regions of the image.
- Page Segmentation: Good OCR software will detect sections of the image that are outside the document.
- Page Layout Analysis: This process classifies documents according to their layouts, such as multicolumn, tables, and graphs, which are segmented into their own blocks.
- Character/Word Segmentation: OCR works best when individual characters are identified and isolated:
- In structuring forms, separating characters such as in the figure above improves OCR performance.

3.4.2.2 Character Recognition

Once the image is preprocessed, pattern recognition techniques are applied to correctly identify the character in the known alphabet. Feature-based techniques rely on explicit aspects of the character such as vertical/horizontal lines, loops, and line intersections. The features of a specific character are compared with a set of features from a known set of characters to identify the most similar character (for example, an "E" is identified from an "F" by three versus two horizontal lines; an "m" is identified from an "n" with a double inverted loop). These techniques are particularly used for handwritten documents where character appearances vary greatly. Typical algorithms include k-nearest neighbors and neural networks.

3.4.2.3 Postprocessing Correction

Even the best OCR program is subject to error, particularly if scanned documents contain handwritten characters. Typical errors include substitution errors (e.g. incorrectly identifying an F for an E), deletion errors, and insertion errors. Processing corrections can be done manually by checking the output flagged by the OCR software as having a high probability of error, or it can be semiautomated by adding a second layer of AI algorithms for a second pass.

3.4.3 Examples of OCR Applications

We illustrate the use of OCR in converting PDF forms to a searchable database by giving two examples. The first converts an exemption form submitted in PDF to a database. The second converts a condominium declaration document.

3.4.3.1 Processing a Submitted Senior Exemption Application

While most jurisdictions have online form submissions, paper applications are still accepted for the segment of the population that does not have computers, has no internet connection, mistrusts online submissions, or is unable for other reasons to submit digital forms.

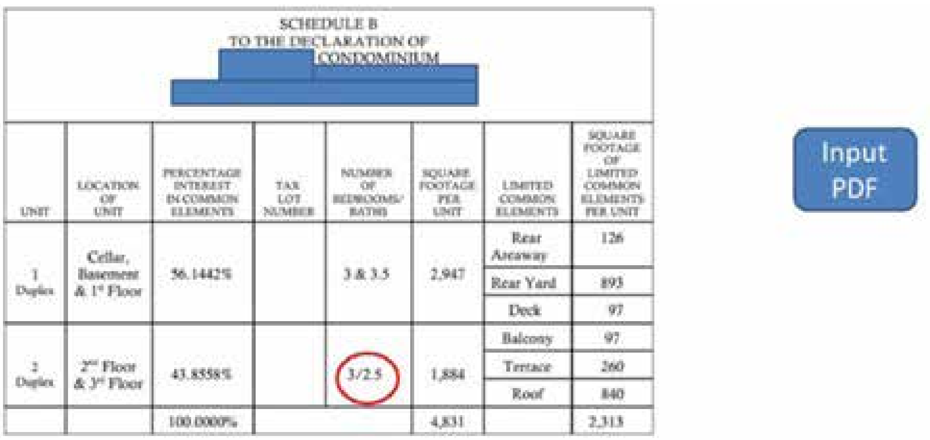
An example of an open-source OCR software is called gImageReader, which can be downloaded from https://sourceforge.net/projects/gimagereader/. The website summarizes its features, which we repeat here. They include the ability to:

- Import PDF documents and images from disk, scanning devices, clipboard, and screenshots
- Process multiple images and documents in one go
- Manual or automatic recognition area definition
- Recognize to plain text or to hOCR documents
- Recognized text displayed directly next to the image
- Post-process the recognized text, including spellchecking
- Generate PDF documents from hOCR documents

The gImageReader is a front end to another widely used open-source OCR software called Tesseract, which uses neural network techniques to recognize characters and correct output.

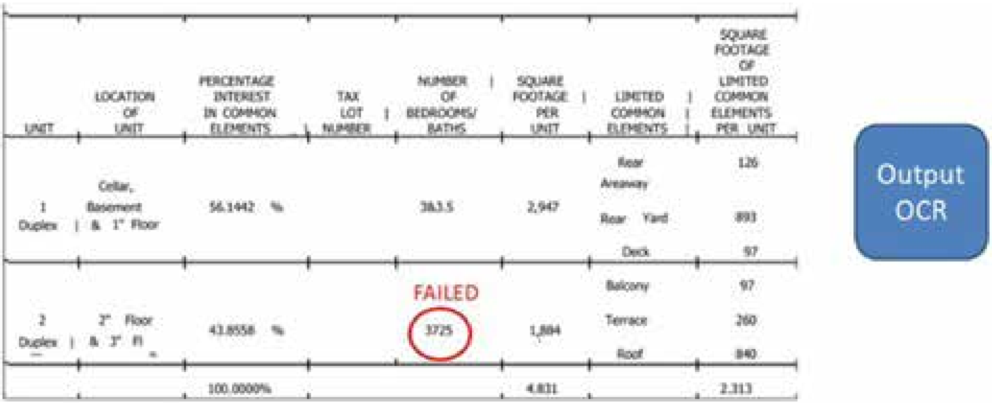
After the paper application is scanned, it is read into gImageReader. The goal is to extract the information the applicant supplied, store it in a database, and down the line merge it with the information of those who filed online. The output provides the text that the OCR reader recognized and the confidence prediction that the text was read correctly. In forms where input is typed, the OCR success is high. For information that is handwritten, the success of the reader varies. In this example, the open-source reader picked up the typed text with over 95% confidence, but the handwritten texts are incorrect, and this is reflected in a confidence level below 80%.

Figure 4. OCR Reader Results


SCHEDULE B
TO THE DECLARATION OF
CONDOMINIUM

UNIT	LOCATION OF UNIT	PERCENTAGE INTEREST IN COMMON ELEMENTS	TAX LOT NUMBER	NUMBER OF BEDROOMS/ BATHS	SQUARE FOOTAGE PER UNIT	LIMITED COMMON ELEMENTS	SQUARE FOOTAGE OF LIMITED COMMON ELEMENTS PER UNIT
1 Duplex	Cellar, Basement & 1st Floor	56.1442%		3 & 3.5	2,947	Rear Areaway	126
						Rear Yard	893
						Deck	97
2 Duplex	2nd Floor & 3rd Floor	43.8558%		3/2.5	1,884	Balcony	97
						Terrace	260
						Roof	840
		100.0000%			4,831		2,313

Figure 5. OCR Result Exported as a Database



UNIT	LOCATION OF UNIT	PERCENTAGE INTEREST IN COMMON ELEMENTS	TAX LOT NUMBER	NUMBER OF BEDROOMS/ BATHS	SQUARE FOOTAGE PER UNIT	LIMITED COMMON ELEMENTS	SQUARE FOOTAGE OF LIMITED COMMON ELEMENTS PER UNIT
1 Duplex	Cellar, Basement & 1st Floor	56.1442 %		383.5	2,947	Rear Areaway Rear Yard Deck	126 893 97
2 Duplex	2nd Floor & 3rd Fl	43.8558 %		3725	1,884	Balcony Terrace Roof	97 260 840
		100.0000%			4,831		2,313

Manual corrections of “Brookly” to “Brooklyn,” “1359” to “1358,” and “=” to “5” are necessary before exporting the result and saving it in a file (Word, Excel, hORC document, or plain text file). This simple example illustrates the importance of vetting the accuracy of the OCR software before purchasing it.

3.4.3.1 Enhancing CAMA Data by Converting Condominium Declaration Documents

When condominiums are formed, their bylaws are recorded together with the building floor plans, unit shares, unit square footage, unit limited common elements, and other physical data that may not be part of the CAMA record. To enhance the CAMA database, the condominium declarations are scanned for “Schedule D,” which, in the jurisdiction used in our example, contains the detailed unit description.

Because these are typed and not handwritten, the accuracy is high. Using open-source software gImageReader, we have one error in re-creating the table. Instead of 3/2.5 baths, the OCR incorrectly outputs 3725. Note however that the OCR output is in a database, so a change of the data from 3725 to 3/2.5 is easily accomplished. The OCR output is not a static image file but a database whose data can be formatted and written into Excel, PDF, Word, etc.

Figure 5 illustrates the usefulness of OCR in transforming documents to a database with high accuracy using free software.

3.4.4 Applications in Property Tax Administration

The main reason an operation would consider OCR is to convert paper information to data through a process that is automated and as accurate as possible. Any workflow with paper that is being entered manually has the potential to be an OCR application. Candidates include:

1. Permit/benefit applications/appeal applications processing – Records that are in physical format whose legacy information is important for future analytics, and submissions that are in paper but have their digital counterparts.
2. Land records – Deeds, agreements, loans, and other forms of conveyances that are in PDF format. The advantage to having land records in a database are many, such as the ability to search by grantor/grantee.
3. Property cards digitization – Government records that are in paper or PDF format.

3.4.5 Conclusion

Assesing jurisdictions are in the process of digitizing records and moving to paperless offices. Scanned information in static PDF or image format is insufficient. Acknowledging the value of data contained in the scanned files results in the need to convert that information to a database as the source for analytics. OCR paired with AI techniques transforms old workflows into a paperless office with data warehousing and analytic capabilities that allow assessing jurisdictions to adapt to the changing needs of the public they serve.

3.5 Three Geospatial Case Studies, Including LiDAR

These three case studies are from a practitioner of GIS and geospatial analytics. They describe the application of AI to geospatial data, increasingly used with digital mapping and analytics in assessment. Geospatial data and GIS software tools are useful for assessment and valuation because they add specific information about location, neighborhoods, physical changes, and geography and can create realistic virtual models of properties, neighborhoods, and cities. Since the data is digital, multivariate, and typically not normally distributed, AI is ideally suited for processing and analyzing that data and is now being used to identify physical features and changes in assessment and valuation. The tax assessment domain is rapidly incorporating more location-aware data throughout its workflows and analytic processes and automating the capture of building physicals has the potential to change valuation and assessment.

3.5.1 Background on Geographic Information Systems and Geospatial Data

Geographic information systems (GIS) are software systems designed around location data that exists within a geographic coordinate system and can be mapped precisely to the Earth's surface. A GIS consists of a database that is spatially enabled, a visualization engine with a coordinate system for mapping, and special algorithms for processing, manipulating, and analyzing spatial data. It is the analytic backbone for using and processing spatial data and enables 3D data methods, flexible scaling, and 3D visualization capabilities. The concept of a coordinate system is critical for a GIS because it creates a model of reality within the software. The GIS profession is one of the early fields utilizing ML and AI because geospatial data sets are large and often multivariate.

GIS capabilities are now being integrated into other types of software-processing, including CAMA systems, which extends their utility and takes them out of the hands of specialists. Because the mapping data is digital and synced to a database, AI and ML algorithms are easily incorporated into GIS software.

Spatial data has inherent geographic, geometric, and spatial information built into it, which gives it characteristics and capabilities that are very useful for assessment and valuation:

- Spatial data has mathematical properties called topology. Topology identifies relationships between entities such as within, outside, adjacent, and intersect.
- Spatial data is good at integrating disparate data types, so it is well-suited for the modeling and visualization of complicated 3D systems for such areas as urban infrastructure and housing. The parcels and buildings that assessors are evaluating can be mimicked in a digital format.
- This data type is not normally distributed in a statistical sense, so standard statistical tools (e.g. regression) are often not applicable.

A geospatial data set can be interchangeably called a layer, a feature class, or an attribute and can be stored in a folder or a spatial-enabled database. There are two general types of geospatial data: vector data and raster data. Vector data is especially beneficial for modeling systems of parcels and land records where exact positions are critical. Raster data consists of a matrix of pixels spread continuously across a surface with a known starting point — it looks like a digital image. Light detection and ranging (LiDAR) are vector data that is sometimes placed in its own data category called "point clouds" because it is collected by a sensor and consists of a huge cloud of points (e.g. billions), which is a critical elevation data set for measuring the shape, geometry and height of buildings and ground surface.

A typical way of organizing geospatial analysis is to separate data into supervised and unsupervised learning tasks. Supervised learning requires prior knowledge, samples, or ground truthing about the systems' outputs. The goal is to link the source geospatial data, like bands of imagery, to known processes and to output some results that match the known output. The function creates a prediction of the output class or variable. Unsupervised methods, also called data mining (see below), seek the inherent structure within the data and do not use explicit labels or samples. These methods self-organize information.

Working with geospatial data allows users to blend a variety of data types such as satellite imagery and LiDAR data with building, neighborhood, and transportation data. This data is processed with a variety of analytic, geospatial, statistical, computer science, and ML algorithms, including some of the methods discussed in the next section. They all follow a similar structure and format: Data is acquired and preprocessed. Training data is created with information tags. An algorithm is applied using unsupervised (e.g. no training data) or supervised (e.g. uses selected samples) approaches. Finally, the validity of the data is established through a quality control and quality assurance process. In the following section, we have three typical geospatial analytic workflows that describe different types of analyses using AI and geospatial data.

3.5.2 Three Case Studies

3.5.2.1 Land-Use Classification

A standard geospatial problem is identifying the locations of a particular type of land use or land cover in a community. Wetlands data, for instance, is often out of date or mapped at insufficiently fine scales to be used effectively by assessors. Wetlands may be undevelopable or have development restrictions, such as buffers or upland review areas, and are important because of their impact on valuation and assessment. A geospatial workflow to identify freshwater wetlands combines ground truth field data, a supervised classification process on imagery, and geospatial processing. A standard data input for this analysis is high-resolution imagery with three visible bands and a fourth infrared band, plus complementary environmental information such as elevation, slope, and land cover. The infrared band is particularly useful for differentiating water and land areas. The analyst first creates a sample of known wetlands areas. Typically, this is a combination of preexisting data from state, provincial, and federal agencies, plus drawn polygons of wetland locations and possible field verified data. These polygons identify areas of pixels in the map that overlap known locations of freshwater wetlands. The wetland polygons or training sites should be exclusively wetlands so that the sample pixels represent the desired class. These identified wetlands have a characteristic spectral signature because of the presence of water and vegetation.

Finally, the data and the samples are run through a supervised training algorithm to link the spectral and environmental characteristics of the geospatial data (imagery, elevation, slope, etc.) with known wetland samples. The statistical methods involved in classification could include a variety of AI approaches, such as random forests, support vector machines, and recently, neural networks/deep learning, which you will learn more about later. Each might be sufficiently effective. A review of literature would help guide your selection. Because the data analysis creates probabilistic, not deterministic, results, a quality assurance process is critical to identify how successful the method is. A "confusion matrix" is used to determine the errors, which can come from sampling, imagery, or processing steps. After the initial quality assurance steps, additional iterations might add rules to improve the accuracy of the model. An example of such a rule is: Wetlands should be only in areas with slopes less than 1% and have a concave shape. To summarize the steps: acquire training samples, find spectral signatures, process the data using AI, then validate results.

3.5.2.2 Change Detection of Building Footprints and Square Footage Calculation Using LiDAR

Another typical problem for an assessment office is determining the gross or total square footage of a building. It is a critical variable for valuation, and assessors have traditionally relied on manual methods such as architectural drawings and hand measurements done in the field. To identify changes in the square footage of a building requires elevation data from two dates, called Time 1 and Time 2. An analyst would start with a baseline in Time 1 and then look for horizontal and vertical changes in the building in Time 2. This analysis could be automated with a method using LiDAR. LiDAR, as already mentioned, is a point cloud data set that is remotely sensed. A laser is typically shot down from a fixed wing airplane and then the reflection returns to a sensor, giving a highly accurate distance of the object that it bounced of off — a building, a tree, or the ground. A LiDAR data set captures one to 20+ points per square meter. It consists of x, y, and z points that have additional information, such as the number of returns, which helps with identifying whether a point is related to a tree, rooftop, or ground plane.

Using LiDAR processing, the data can be split into two surfaces through a process of interpolation where points of LiDAR are turned into a continuous surface model in a raster data format. One raster surface is called the digital elevation model (DEM), which defines the ground plane. The other raster surface is called the digital surface model (DSM), which defines the surfaces above the ground plane, including rooftops, trees, and wires. Within the defined footprint of each building acquired from GIS data, the two surfaces can be compared with and subtracted from one another so that at each raster pixel within the footprint we know the height from the ground plane to the roof of the building. The aggregated difference defines the volume of the building.

With two dates of LiDAR the volume of one building can be compared against the other. If a volume increases in the second year, then the building can be flagged for inspection. The total square footage can then be estimated by taking the volume divided by the number of floors. This method can be highly accurate for buildings with standard floor configurations but is susceptible to errors from tree canopies hanging over buildings and different specifications for the two compared data sets. As is typical of geospatial and remote sensing analytic workflows, a variety of algorithms are used, including geospatial processing, interpolation and statistical methods, and AI methods to clean, subset, classify, and analyze the data.

3.5.2.3 Building Feature Detection

A typical and simple problem in an assessment office is identifying physical features in buildings, such as external stairs. The absence or presence of a staircase in front of a residential building in a city like New York can be used to help identify whether the building has an occupied basement. These physical features are often underreported and may be illegal, depending on local ordinances. The assessor might conduct a field survey for this feature, but this would require many hours of visits. An automated method will screen the larger pool of data and allow the assessor to focus on flagged cases.

The assessor could use convolutional neural networks (CNN) to make a feature identification model to predict the absence or presence of stairs. CNNs (discussed in more detail later) are well-suited to identifying items within image files, specifically the absence or presence of a feature. Surprisingly, each pixel within a digital image can function as an input variable for a CNN, so to simplify the data, "pooling" is used. CNN is unique because a hierarchy of layers (e.g. a deep model) allows the physical structure of the image, such as the location of edges, to be added as part of the input. The software required to run a CNN model is available in the open-source Python environment — freely available for anyone to download.

Back to our example: The first step would be to "train" the CNN using supervised learning. We would use digital images of building facades taken from an administrative field inspection and, after visually inspecting them, tag those images that have stairways and those that do not. Accuracy is very important in this step because this training information creates the optimized function. The goal of the training data set is to include a variety of conditions that might impact the accuracy of the model, such as tree shadows and fencing in front of the buildings. With these inputs we conduct "parameter" tuning to check the model for stability and overfitting. We then run the model on data that was not part of the training data, called the hold-out data set. In a pilot study in New York City, a model was 90 % accurate at identifying buildings with external stairways in field photos taken by assessors. Once a model is created and validated, it can be placed in an automated workflow.

3.6 AI Search and Optimization in Action: "AutoML"

In this case study, the utility and power of combining computer science-based search methods (see the Methods section) and ML is discussed. To stress the critical importance of AI search and optimization techniques to appraisal, we present the results of a small experiment with automated machine learning (AutoML) that demonstrate the capabilities of ML techniques. Because AutoML automates the creation of ML models for nonexperts, it is a powerful and useful tool. AutoML implementations use the types of search and optimization techniques discussed later to automatically engineer features, select the best types of ML models to use, and calibrate and combine the selected model types.

This experiment sets out to see how well AutoML search/optimization procedures compare with those of a skilled MRA modeler and to see whether the AutoML reduced the effort of the modeler. The experiment is from a single neighborhood with 1,560 sales that was used to calibrate both an MRA and a comparable sales model (Borst 2020). The performance of those traditional valuation approaches is shown in the first two rows of *Table 1:* (1) CAMA Comparable Sales and (2) CAMA MRA.

Initially, the AutoML search and optimization procedure was provided the same data, including the hand-engineered features from the first two approaches. Because of the holdout and verification data processes, the number of sales provided to (3) AutoML base data w/CAMA Analyst transforms was less (only 322). Despite the reduced training data, AutoML produced comparable results in row 3.

A second AutoML calibration was performed in row 4, (4) AutoML base data w/auto transforms only. However, for this experiment, the analyst's handcrafted transforms (features) were removed, and the AutoML procedure no longer had the added advantage of curated features. Even with only the raw data, row 4 results also show that AutoML produced results comparable to those of the expert modeler.

Table 1. Compares search/optimization based AutoML with that of traditional manual CAMA MRA and comparable sales approaches

Method	Count	Median	WgtMean	COD	COV	PRD	PRB
(1) CAMA comparable sales	1,560	1.00	1.00	3.78	4.97	1.00	-0.04
(2) CAMA MRA	1,560	1.00	1.00	4.16	5.32	1.00	-0.04
(3) AutoML base data w/CAMA Analyst transforms	322	1.00	1.00	3.98	5.49	1.00	-0.05
(4) AutoML base data w/auto transforms only	318	1.00	1.00	3.94	5.20	1.00	-0.01

Borst shows that the AI search and optimization approach using AutoML to automate market modeling is competitive with the traditional approaches to MRA modeling for this single experiment. This result is remarkable in such a simple experiment. AutoML simplified the task of engineering features and avoided the countless manual trial runs of possible ML architectures and hyperparameter settings. The laborious work shifted from the modeler to the AutoML software.

3.6.1 Discussion and Conclusion

As with ML techniques, there is no single best AI search or optimization technique. The selection of the most appropriate AI technique often depends on the structure of the problem, constraints, and the objective function form.

Most ML procedures use a small search step in their training iterations. As a result, most ML "weight learning" procedures are directly cast and solved as a search and optimization problem such that the error function against the training data is minimized. ML procedures solve classic optimizations using an iterative method that heavily relies on feedback. Feedback is the error information "observed" and used to steer weight space or regression tree searches to improve the model in subsequent iterations. Many, if not most, of the ML techniques are subject to the risks and limitations of local maxima and minima.

ML's hyperparameter optimization processes are generally search and optimization procedures as well; the search and optimization procedures discussed in this section are all relevant and directly useful today for hyperparameter tuning across the various ML techniques. We also see that AI search and optimization techniques can be combined into powerful AutoML approaches that automatically synthesize features and models that preliminary experiments have shown to be at least comparable with MRA models handcrafted by industry experts.

AI search and optimization can be used by today's assessment offices and IT practitioners to greatly improve the performance and power of hearing scheduling, inspections scheduling and routing, and comparable property selection found in the appraisal office. We feel this latter area of business process optimization and decision optimization, rather than the narrower confines of valuation model calibration, is the best path to create business value in the assessment office.

3.7 Use of Technology and AI in Africa: Case Study Examples

In this final case study, a brief review of assessment solutions in Africa demonstrates the utility of integrated geospatial data, imagery, and AI. These countries are geographically large and lack wellfunded governance; however, they are using integrated digitalization to move rapidly forward. The full digitalization of their assessment and valuation workflows combined with the use of imagery allows them to modernize their systems without going through and updating many of the legacy operational processes found in better established or wealthier tax assessment organizations. They are combining AI and ML workflows that integrate geospatial data and cadastral mapping similar to that which was reported in the geospatial case study.

In Africa, the application of digital technology and data has been of fundamental importance for revolutionizing property taxation, though the technical complexity of ensuring that tax rolls are complete and valuations current is often perceived as a major barrier to property tax collection across most developing nations. The identification of taxable properties is often inadequate or incomplete, out of date, and fragmented between government units (Ali et al., 2018) and it is not uncommon to find fewer than half of taxable properties to be on the tax roll (Blochliger, 2015). There has also been a general lack of political will, inadequate records of taxable properties, and a lack of skilled personnel, which also limits operational efficiencies (Balogun, 2019; Muhammad and Ishiyaku, 2013; Oluwadare and Ojo, 2014). Consequently, several African nations have turned to data digitization, digitalization, and image processing (Olatunji and Ayodele, 2017), as well as other AI practices and techniques, and combined them with GIS to transform various aspects of the property tax system.

As discussed previously, powerful analytic capabilities are unleashed when geospatial data is combined with AI. There are numerous application and cutting-edge approaches and technologies, such as detecting changes using remotely sensed images like Landsat, and augmenting appraisal accuracy for property tax assessment and administration using georeferenced imagery. Major advances in the design and implementation of image processing with satellite imagery allow for efficient cadastral management. Equally, advances in technology and AI have increased administrative and enforcement efforts through digitization and digitalization.

3.7.1 Examples of Innovative Practices in Four African Countries

Below are key examples of how Rwanda, Nigeria, Uganda, and Zambia have used and integrated AI to improve their taxation systems.

Rwanda is the only African country that has established a complete and fully digital legal cadastral database — the land administration information system (LAIS) that contains spatial and textual data for all registered land parcels. Despite this, Rwanda's property registry does not contain the level of detail necessary for mass valuation exercises. Therefore, recent use has been made of remotely sensed high-resolution satellite imagery to enhance data digitalization and to create tax maps for revenue collection. The use of AI through data digitization has permitted the automation of segmentation techniques to extract building footprints and building heights, which provides increased coverage for tax rolls and change detection for estimating property taxation values. This nuanced data in conjunction with spatial data routinely available from statistical institutes is helping inform routine processes of CAMA.

> *In Nigeria, numerous exercises have been undertaken using remote sensing, specifically satellite imagery and digital image processing techniques that include image enhancement and filtering. These practices have helped improve image quality. And using AI image classification facilitates the creation of land use and land cover layers to be integrated within GIS, which forms the basis for identifying land-use changes and helps the viability of CAMA.*

In the Kampala Capital City Authority (KCCA) of Uganda, recent automation efforts and data digitization and digitalization have helped create a new city-level taxpayer database that has been automated as part of a new electronic revenue management system called eCitie (Blake and Kriticos, 2019). This administrative and collection system automates the generation of an account for each taxpayer and calculates the payments required for each account.

This digitalization enables taxpayers to access their accounts remotely at any time for information and payment and make remote payments through a number of convenient and traceable channels It also facilitates automation of receipts for payment, sends out reminders to taxpayers, flags which taxpayers are in arrears, and generates management reports for the assessment office (Blake and Kriticos, 2019). From an enforcement perspective, this digitalization also helps prioritize and identify revenue sources and taxpayers who are the most significant. The rollout of this automated system, and its successful design and operation, have required a critical mass of qualified staff, who receive intensive and periodic training and continuous professional development on taxation policy, laws, and operational guidelines for the technology.

In Zambia, they are incorporating mobile mapping tools with GIS. They developed a pilot mobile GIS using open-source software tools and mobile cloud computing (mapping services) to digitize property maps. This allows them to incorporate real-time capture of attributes with spatial and image data of properties for the development and implementation of the mobile mapping applications (Neene and Kabemba, 2017). This digitization workflow determines the satellite image of the property mapping area to create the land parcel, which then saves GPS coordinates to the cloud database. This digitalization process provides a mapping system sequence of classes modeled on the map, location, land, owner, property, valuation, and individual entities in the property mapping problem and provides an efficient and useful tool to map properties in the real-time setting.

4. METHODS

4.1 Machine Learning Methods

4.1.2 Motivating Examples

In this section, we review typical methods used ML and introduce the term data mining. As a reminder, ML is a subfield of AI and it is a set of computer algorithms that learn from training data to make predictions without explicit programming. We provide examples of possible use cases for ML methods and models that can be applied by assessment jurisdictions. These are commonly used techniques in data science and geospatial analytics. The section is meant as an introduction that will lead the reader into further investigation on the more technical subject matter.

In Table 2, we start with a list of common analytic problems (see column 1) that assessors face and In and then possible ML solutions (column 2). Hopefully, you will recognize issues that you face on a regular basis.

Table 2. Assessment problems and possible analytic solutions

Analytic Problems (column 1)	Potential ML Solutions (column 2)
The assessor would like to scope out a future reassessment project and is looking for a new method to examine the data and highlight possible data quality issues. They have an idea of the property characteristics they would like to reassess for accuracy — for example, bedroom and bathroom counts — and they would like to be able to describe the relationships between these and other property characteristics and identify potential inaccuracies.	Regression decision tree models may help identify the most important property characteristics for predicting bedroom or bathroom counts, describe the relationships between other property characteristics and bedroom or bathroom counts, and provide indicated predicted values to compare actual values to and highlight differences.
The assessor has performed several reassessment projects in the past and tracked whether the projects have identified and made changes to individual properties. They would like to gain insight into factors that may influence the capture of new or updated inventory.	A classification decision tree could be designed with a binary flag as the response variable indicating whether a reassessment project has captured new or updated inventory in the past. This decision tree model may help identify property characteristics of importance to a reassessment project. On a future set of properties, the model could be applied to predict the probability that a given property will have new or updated inventory captured.
The assessor has existing multiple linear regression models that predict market value of properties. They would like to examine whether there are other ML models that can predict market value with greater accuracy.	A random forest or gradient boosting decision tree model could be applied to predict market value and compare results.
The assessor would like to reexamine how market sets are defined in a particular area. Multiple clustering models could be built, each presenting different numbers of clusters, using important local and market characteristics.	Local subject matter expertise could be applied to the multiple cluster models to select the number of clusters that best describes the market and meets the needs of further analysis.

4.1.2 Data Mining and Machine Learning

Data mining is the process of discovering patterns in and extracting useful knowledge from data sets and transforming this information to provide further insight. It is the combination of multiple disciplines, including statistics, ML, database technology, pattern recognition, AI, information retrieval, network science, high-performing computing, and data visualization (Han et al., 2012).

Data mining is used to discover a variety of patterns in data sets, either by providing insight into characteristics of a known class or grouping of data, or by seeking to describe the classes or concepts within a data set when they are previously unknown. It is especially valuable for data with complex behavior or with many variables. Within a data set with known classes or groupings, data mining can summarize characteristics, compare characteristics of multiple known classes of data, identify correlated or associated characteristics, and detect and analyze outliers (Han et al.2012).

While data mining and ML often use the same methods, data mining is primarily focused on discovering previously unknown characteristics of a data set, whereas ML is primarily focused on learning how to make sense of a data set so that predictions can be made on a new data set. Data mining may even be used on a data set before an ML method is applied to improve the accuracy of the ML model.

Within a data set with unknown classes or concepts, ML methods can perform classification (categorical class labels) or regression (numeric data values) based on the analysis of an existing training data set. We refer to this also as supervised learning. Within a data set with unknown classes where there are no preexisting labels and no known training data set, ML methods can perform clustering to determine these classes. This form of

ML, where previously unknown classes and values are extracted from data sets, is further described in this section and is also called unsupervised learning.

There are many ML methods to perform classification, regression, or clustering. These include, but are not limited to:

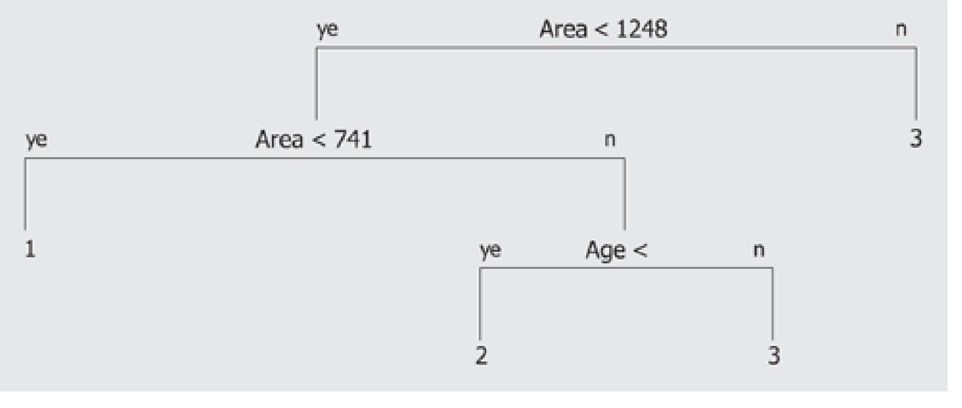
- Neural networks (discussed in a separate section)
- Decision trees
- Clustering algorithms
- The k-nearest neighbor algorithm

The following sections describe in more detail several decision tree methods, the k-means clustering algorithm, and the k-nearest neighbor classification algorithm.

4.1.3 Tree Methods

Decision trees are one of the most popular techniques used in data mining, particularly for classification (Gorunescu, 2011). A decision tree is a predictive model that is a series of one or more "if-then" statements that can be used for either regression or classification. The if-then statements segment the predictor variable into two or more simple, nonoverlapping regions used to predict an outcome. A decision tree is generally fast to construct, has interpretable and easy to visually interpret results (provided the tree is small), and can be built from a variety of data types, including numeric and categorical predictor variables, even if there are missing values. Decision trees are also not influenced by outlier values and are generally minimally impacted by the inclusion of many irrelevant predictor variables (Hastie et al., 2009).

Figure 6. An example decision tree predicting the number of bedrooms in a townhouse



The if-then statements of a decision tree segment the predictor variables into several simple, nonoverlapping regions. Within each of these simple regions, a model is used to predict the outcome. A decision tree gets its name because the series of if-then statements that segment the predictor variables and arrive at predictions can be summarized in a tree. The decision tree (Figure 6) shows a theoretical, simple prediction of the number of bedrooms in a town house. This decision tree could also be described as:

- if Area < 1,248 sq ft and Area < 741 sq ft then Bedrooms = 1
- if Area < 1,248 sq ft and Area >= 741 sq ft and Age < 44 then Bedrooms = 2
- if Area < 1,248 sq ft and Area >= 741 sq ft and Age >= 44 then Bedrooms = 3
- if Area >= 1,248 sq ft then Bedrooms = 3

A decision tree is described as having nodes and branches. Terminal nodes, or leaves, of a decision tree are the ends of the if-then statements. In the example tree, the terminal nodes are the number of bedrooms predicted by the tree. The points along the tree where a split occurs are called internal nodes. In the example tree, the internal nodes are indicated by "Area < 1248 sq ft," "Area < 741 sq ft," and "Age < 44." The branches of a decision tree are the segments that connect the nodes.

A regression tree predicts the same numeric response value for every observation that falls into a given region. It does so by taking the mean of the response values for all observations in the training data set within the region. The tree is built by taking an approach known as recursive binary splitting, which can be described as a top-down and greedy (e.g. the most optimal) approach: starting at the top of the tree where all observations are in one region, the tree is split based on the best split at that particular step, rather than looking ahead to the split that will lead to a better tree (James et al., 2013). The best split is defined as the one that minimizes the residual sum of squares (RSS). Each branch is subsequently split using the same top-down, greedy approach until a specified stopping criterion is reached; for example, that each terminal node has too few observations to be split further.

A classification tree similarly predicts the same categorical response for every observation that falls into a given region, though in the case of classification the response is a categorical class label. Whereas a regression tree takes the mean of the response values for all observations in the training data set within the region in question, a classification takes the mode, or most commonly occurring, class. A classification tree is built using the recursive binary splitting described for regression trees; however, rather than using the RSS, a classification tree uses splitting criteria, which minimizes the "impurity" of a node (James et al., 2013). The purer a node, the fewer observations in the training data set that disagree with the chosen class. The possible measures of node purity are called the misclassification error, the Gini index, or cross-entropy (Hastie et al., 2009).

A decision tree is likely to create a complex series of splits that produces good predictions on the training data set but is likely to overfit in the process and produce poor predictions on a test data set. To reduce this variance and to increase the interpretability, a decision tree is often grown to a large size and then pruned back to a smaller subtree. To find the best-sized tree, an error rate that penalizes for the complexity and number of terminal nodes is used in conjunction with either a validation data set or cross-validation to find the pruned tree with the lowest penalized error rate (Kuhn and Johnson, 2013, 177-8).

A decision tree is a particularly useful ML technique for assessment and valuation because it is easily interpretable and easily visualized. It additionally performs feature selection, which can be helpful to identify those predictor variables that are influential in predicting the value or class response. However, a decision tree predicts only a limited number of response values or classes; a tree can predict at most the number of distinct response values or classes as number of terminal nodes. Additionally, tree models partition the data into rectangular regions, but if the relationship between the predictor variables and response is not adequately described by these partitions, then the predictive accuracy will suffer (Kuhn and Johnson, 2013). For these reasons, a decision tree typically has worse predictive accuracy compared with other approaches. However, predictive accuracy generally improves with more complex tree methods where multiple trees are aggregated to yield a single consensus prediction (Hastie et al., 2009). These complex tree methods include bagging, random forests, and boosting.

4.1.3.1 Bagging

Bagging, which is short for bootstrap aggregation, is an ensemble procedure (using multiple data sets) applicable generally to ML algorithms to reduce variance and limit overfitting (James et al., 2013). For these reasons, bagging is often applied to decision trees to improve their predictive accuracy, as they generally have high variance and a tendency to overfit to training data. The concept behind bagging is that the variance of an averaged set of n observations, each with its own variance, is proportional to 1/n. Therefore, increasing the number of observations that are averaged decreases the variance of the average.

In terms of decision trees, this means that variance can be reduced by building multiple trees from distinct training sets drawn from the population and averaging the multiple resulting predictions. In practice, it is generally not possible to make multiple training data sets, so instead bootstrapping is applied. That is, multiple data sets are drawn from the training data set using random sampling with replacement to create several bootstrapped training data sets. A decision tree is then built on each of the bootstrapped training data sets, and the resulting predictions from the multiple decision trees are averaged to obtain a final aggregate prediction. Hence, bootstrap aggregation, or bagging. When bagging is performed on a regression decision tree, the final aggregated prediction is simply the average, as described above. For a classification decision tree, the class predicted most often among the bootstrapped trees for each observation is selected as the final aggregated prediction.

In addition to reducing variance when building decision trees, bagging also has the advantage of a built-in method of testing the error of the bagged model, known as the out-of-bag (OOB) error estimate (Kuhn and Johnson, 2013). For each bootstrap sample drawn to build the bagged model, certain observations are left out of the bootstrap sample by the nature of resampling with replacement. It is possible to fit each of the bagged models to those observations that are OOB and aggregate the OOB predictions for each observation. In this way, every observation will have a consensus OOB prediction, from which the regression or classification error can be calculated.

A further advantage of bagging decision trees is that increasing the number of bootstrapped training data sets does not lead to overfitting (James et al. 2013, 317). In conjunction with the OOB error, the number of bootstrapped data sets can increase to a sufficiently large value until the OOB error no longer decreases meaningfully.

However, while bagging has many advantages for decision trees, it does reduce the interpretability of the resulting predictions. Whereas a single decision tree can be easily described with a tree visualization or a series of if-then statements, a bagged decision tree that results in aggregated predictions cannot be described as clearly. It is, however, possible to describe the importance of each variable in the final bagged decision tree.

4.1.3.2 Random Forests

A random forest is a model where decision trees are built on multiple bootstrapped data sets drawn from a training data set, with the multiple resulting predictions aggregated to reach a final prediction. Unlike bagging, each split is allowed to consider only a random subset of predictor variables. A random forest model is similar to a bagged decision tree in that it also uses several decision trees built on bootstrapped training data sets and is widely used with remote sensing and geospatial data. However, unlike bagging, which considers all predictors when building the multiple decision trees, a random forest is restricted so that at each split in every decision tree built, only a random subset of the predictors can be considered for the split. That is, if a decision tree has p total predictors, at each potential split in a random forest, a subset m of predictors is randomly selected as the candidate predictors for the split where $m < p$ (note $m = p$ is the case for bagging). Typically, the number of predictors selected as candidates for each split is approximately the square root of the total number of predictors (m □ □p) (Hastie et al. 2009, 589). Each tree built in this way is unpruned and grown to its maximum size (Han et al., 2012).

A random forest model may provide better performance and reduce variance over a bagged model if the bagged trees are highly correlated. That is, if a data set has one variable that is a strong predictor of the response and other variables that are moderate predictors of the response, a decision tree built on one bootstrapped data set is likely to select the strong predictor in the top split, and so are any other trees built on other bootstrapped data sets (James et al., 2013). What this means is that a set of bootstrapped decision trees in this case would be very similarly built, with a similar sequence of if-then splits defined for each of the bootstrapped trees. In an instance such as this, taking the average of multiple correlated trees will not greatly reduce the variance (Hastie et al., 2009).

By considering only a randomly selected subset of predictors at each split in a bootstrapped decision tree, several trees will be built that do not select the strong predictor as the first split; the strong predictor may not even be considered for any split. By restricting the number of predictors at each split and adding this randomness, a random forest model can build several bootstrapped trees that are uncorrelated and therefore reduce the overall variance of the averaged final model. As with bagging, increasing the number of bootstrapped training data sets for random forests does not lead to overfitting, and the number of bootstrapped data sets can increase to a sufficiently large value until the OOB error no longer decreases meaningfully (Kuhn and Johnson, 2013).

4.1.3.3 Gradient Boosting

Like bagging, boosting is a general procedure that can be applied to different ML algorithms; however, in this case it is being described in relation to decision trees. Unlike bagging, which grows trees on multiple bootstrapped data sets and averages the trees, boosting involves growing trees on a modified version of the original data set and repeating the process sequentially. The boosting part of this method indicates that it is built sequentially: after an initial tree is built, a new tree is built off the residuals of the first tree (rather than the outcome variable), the current model is then added to the previous model and the residuals are updated, and the process continues sequentially a specified number of times (James et al., 2013, 321-22). The gradient part of this method indicates that each sequential tree seeks to minimize a function, known as a loss function, that describes how well the predicted values are fit to the true values; for example, minimizing the mean squared error in a regression gradient boosting model (Kuhn and Johnson, 2013). In this way, gradient boosting seeks to add to the current model by focusing the next sequential tree so that it changes predictions to reduce error the most (Elsinghorst, 2018).

One of the tuning parameters for gradient boosting is the number of splits allowed in each tree, which limits the size of each tree built. By fitting small trees to the residuals, the boosting model learns slowly and improves the prediction in areas where is does not perform as well (James et al., 2013). The ideal number of splits allowed for each tree can be determined by testing on a validation data set; it is likely that trees with one to 10 splits allowed will perform well with gradient boosting (Hastie et al.,2009, 363).

Another tuning parameter for a gradient boosting model is the number of trees built. Unlike bagging and random forests, it is possible for a boosting model to overfit to the training data set because each step aims to reduce the error (Kuhn and Johnson, 2013).

Cross-validation can be used to determine the best number of trees built. Another tuning parameter for a gradient boosting model that also helps reduce overfitting is a shrinkage parameter, which is a small positive number between 0 and 1, but typically closer to 0.01 or 0.001 (James et al., 2013, 322). The shrinkage parameter is used so that for each subsequent model built, only a fraction of the subsequent model is added to the previous iteration, where the fraction is the shrinkage parameter. That is, if $\hat{f}(x)$ is the boosting model built to that point, and $\hat{f}^1(x)$ is the next sequential tree model, then the model is updated by $\hat{f}(x)+\lambda\hat{f}^1(x)$ where λ is the shrinkage parameter (Hastie et al., 2009, 364). In this way, the shrinkage parameter helps slow the boosting process down further, with the tradeoff that the number of trees built is likely to increase.

4.1.4 Clustering

Clustering is an unsupervised process of grouping a data set into subgroups, or clusters, so that observations within a cluster are quite similar to one another, while observations between clusters are quite different from one another. Unlike classification, where the class labels are known on a set of training data, in clustering, the groupings are unknown in the training data and need to be learned. Clustering involves defining what makes an observation similar or different and as such is aided by subject matter knowledge of the data set being examined. Clustering may be used in property appraisal to describe market sets, for example, where the clusters may be defined based on characteristics of a building such as size and age.

4.1.4.1 K-means

K-means clustering is a technique in which a data set is split into k distinct, nonoverlapping clusters. That is, for a given number of clusters, k, each observation in a data set belongs to one of the clusters, and no observation belongs to multiple clusters. Observations are assigned to clusters so that the within-cluster variation summed for all clusters is minimized; commonly, the total within-cluster variation is measured as the squared distance between each point and its cluster's center, summed for all observations (Han et al., 2012). Note that the distance between a point and a cluster's center is multidimensional based on the number of variables being considered.

The number of clusters described by k-means clustering is defined by the user. The user-defined number may be directly dictated by a business case; for example, a user may wish to assign properties for inspection to five employees, with the goal of having employees inspect similar properties. In this case, k-means clustering could be applied to a data set with k = 5. However, a business case for k-means clustering may not have a well-defined number of clusters to output. In cases like this, it may be helpful for the user to apply the k-means algorithm multiple times with different numbers of clusters included. For each different value of k, the user could then examine the characteristics that distinguish the clusters and select the number of clusters that is the most interpretable or most useful. For example, if k-means clustering is used to describe market sets, a user may try k between three and seven and select the value of k that provides enough differentiation between the important property characteristics, but not so much differentiation that less meaningful characteristics are described by the clustering.

Computationally, considering every possible way to partition a set of observations into k clusters is very taxing; for this reason, rather than considering the best solution globally, the k-means clustering algorithm seeks to find a local minimum for the within-cluster variation (James et al., 2013). The k-means clustering algorithm begins by randomly assigning k initial cluster centroids. The following two steps are then iterated until the cluster assignments stop changing (Hastie et al., 2009, 460):

- Identify the subset of observations that are closer to a given cluster center than any other and assign the observations to that given cluster. Do this for all cluster centers.
- For each cluster, the mean value of each included variable is computed based on the observations assigned to that cluster. The mean value vector (composed of the means of each included variable) computed for each cluster now represents the cluster centers.

Once the cluster assignments have stopped changing after repeating the above steps, we can consider it a local solution to minimize the within-cluster variation. However, as mentioned previously, to make the k-means algorithm computationally feasible, the cluster definition arrived at is dependent on the initial random assignments. James et al. (2013, 388-389) suggest running the k-means algorithm multiple times with different initial random assignments and selecting the final clustering that minimizes the within-cluster variance.

4.1.5 K-Nearest Neighbor

K-nearest neighbor is an algorithm that can be used for classification and regression. It is widely used for pattern recognition (Han et al., 2012) and has been successfully applied in many classification problems such as handwriting examples and satellite image scenes (Hastie et al., 2009). The basis of the k-nearest neighbor algorithm is a comparison between a test observation and training observations that are like it. For a given test observation, the training data set patterns are searched, and the k training observations that are closest to the test observation are used to compute either the classification or regression value of the test observation. For classification, the most commonly occurring class in the k-nearest neighbors is used to define the class of the test observation, while for regression the average value of the response variable on the k-nearest neighbors defines the response for the test observation.

The closeness of the training data set to test observations is measured using a distance metric; several distance metrics are options, but the simplest and most commonly used is the ordinary straight-line distance, known as the Euclidean distance (Kuhn and Johnson, 2013). Since the features of the data set may be on quite varying scales (for example, comparing the size of a house to the number of bathrooms), it is typical to normalize the features so that they are all on the same scale, for example by standardizing (mean 0 and variance 1) or using min-max normalization (scales to a range from 0 to 1). The distance metric inherently assigns equal weights to each included predictor variable; however, some of these variables may include more noise or be less relevant than other variables, which may impact the accuracy of the k-nearest neighbor algorithm. To account for this, attribute weighting and attribute pruning for noisy variables can be included in the algorithm (Han et al., 2012, 425).

The value of k to use can be determined by examining either the test error or cross-validation error on a set of different k values and choosing the one that minimizes the unseen error. K = 1 often has low bias but high variance, with a decision boundary that is too flexible (Hastie et al. 2009, 465-8). A large value of k becomes too inflexible and produces a decision boundary that is closer to linear, with a low variance but high bias (James et al., 2013, 40-42).

4.1.6 Machine Learning in Practice

In practice, there are many possible situations where applying one or more of the ML methods described above can help discover patterns and extract useful knowledge from raw data, such as the examples at the beginning of this section. The following are questions that may help highlight areas where ML methods could be applied. This is not an exhaustive list, as there are any number of situations that may benefit from one of the classification, regression, or clustering ML models described here.

- **Question**: Is there a business problem where you would like more data-based information to provide insight that can be formed as a classification or regression problem?
    - A decision tree-based or k-nearest neighbor model may be considered in addition to or instead of a more traditional logistic regression or multiple linear regression model.
- **Question**: Are there processes where identifying outliers can provide more useful information?
    - A clustering or k-nearest neighbor algorithm may be employed to define boundaries beyond which a test data point can be considered an anomaly or outlier.
- **Question**: Does your data set include a heterogeneous mix of data types, including nonnumeric, categorical data? Does your data have nonlinear relationships that are difficult to describe using linear regression?
    - A decision tree-based model can handle nonnumeric data in addition to numeric data. It also does not assume linear interactions between variables.
- **Question**: Would you like to understand which variables are most important to a response variable?
    - Decision tree-based methods naturally select the most important features in a regression or classification model and can quantify the importance of each variable within the decision tree.

4.1.6.1 Data Sets

There are no specific data sets to be used for the data mining and ML methods described here; what data is included in a model depends on the business problem the model seeks to inform, though it can include structured, unstructured, and geospatial data. The following are possible considerations for selecting a data set to use with one (or more) of the ML methods described in this section.

- A classification or regression decision tree model can be applied to the same data set that has historically been used with logistic regression or multiple linear regression.
- A decision tree model has naturally built-in feature selection; a data set with many predictor variables may be used successfully without the need to perform feature selection in advance of running models. However, note that variables that are informational, such as ID variables, should be excluded from a decision tree model, as they may otherwise be selected.
- The naturally built-in feature selection of decision tree models provides a means of describing the importance of variables. This may subsequently help inform data sets for other models, such as a multiple linear regression model.
- Decision tree models can include nonnumeric variables. Additionally, these models do not assume linear relationships. Therefore, these models do not require transforming ordinal or categorical variables prior to applying the models, and ordinal variables do not have to have a linearized relationship explicitly defined.

For example, the relationship between a "poor" quality and a "good" quality building does not have to be explicitly defined before applying the model.

- K-means clustering, and k-nearest neighbor models are more dependent on the number of included predictor variables than decision tree models and are therefore also more dependent on the selection of which variables to include. Both of these models can suffer from what is known as the "curse of dimensionality" — that is, as the number of variables included increases, the distance measures used to describe within-cluster variation (for

k-means clustering) or closeness between a test point and training data set (for k-nearest neighbor) converges so that there is little distinction in any distance measure between points (Google, 2020).

- As both k-means clustering and k-nearest neighbor are dependent on distance measurements, both are dependent on the scale of the variables included. Therefore, data sets used for either of these methods will usually have a method of standardization applied (e.g. standardization for mean = 0 and variance = 1 or min-max normalization).

4.1.6.2 Software Tools

There is no one tool or software that is best at carrying out the ML methods described here, and many of these techniques are being incorporated into enterprise software packages. Choice of software may depend on the user's comfort and familiarity with certain tools and the specific algorithms and additional analysis that the user may wish to apply, among other considerations. Examples of software that can be used for data mining and ML include, but are by no means limited to:

- IBM SPSS Modeler
- SAS Factory Miner
- R programming language
- Python programming language
- RapidMiner
- ArcGIS Pro
- SPSS
- Stata

4.2 Artificial Neural Networks and Deep Learning

Artificial neural networks (ANN) are powerful analytic tools that allow the optimization of functions from training data and that have been found to be particularly useful in extracting feature information from images. Conceptually, a simple neural network (NN) consists of an input layer, a function layer(s), and outputs. The popular press and public invoke the term AI when they think of consumer-facing features in smartphones, such as speech recognition, but these features really come from a more complex form of neural network called deep learning. DL is a class of ANN characterized by more layers and more neurons that extract more complicated interactions between variables, such as higher-level features like the shapes of buildings. This section devotes attention to these models because of their flexibility in terms of data types, their adaptability to problem types, and their utility for image classification and pattern extraction. They may prove particularly useful to automate the identification of physical change in buildings.

4.2.1 History

We can break the history of DL into three periods. The early work, before 1970, created NNs that predated AI and used the simplest neural network architecture — the perceptron. NNs in this first wave of research were simple linear models called cybernetics (Russell and Norvig, 2021; and others). Research in the area largely died off until the 1980s, when AI research moved into the world of encoding symbolic logic, embodied in expert systems, and other activities where even the limited computing resources of the day excelled relative to human abilities. The second wave of research in ANN, known as connectionism or parallel distributed processing, brought significant advances toward enabling DL today. The core premise was that many simple functions working together can represent complicated functions (Goodfellow et al., 2016).

The third wave of research and commercial interest in ANNs, the DL era began in 2006 and exploded in 2012, when Geoffrey Hinton and two of his graduate students presented the first neural network, called AlexNet, that dramatically outperformed classical ML methods in image classification (Goodfellow et al., 2016; and others). Since then, there have been advances in deep neural network architectures specific to:

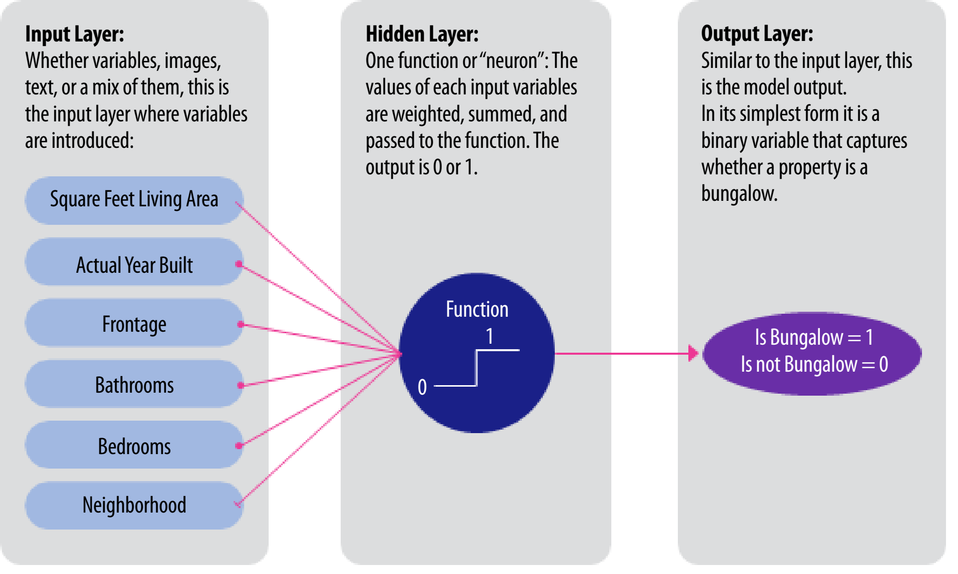
- Computer vision, object identification, and image extraction using convolutional neural networks (CNN)
- Robotic controls and learning machines in the form of deep reinforcement learning
- Signal processing for speech recognition and time-series methods in the form of recurrent neural networks (RNN)
- Language models for translation, chat, and auto-completion
- Generative adversarial networks (GAD) with wide-ranging applications, but most widely known for deep fakes

The power ANNs bring stems from their functional flexibility. Unlike classical ML techniques, where the modeler is responsible for creating the variables that capture things such as diminishing returns compared with the number of bedrooms (e.g. the fifth bedroom doesn't contribute the same amount of value as the third bedroom), DL gives the model the ability to learn those nonlinear relationships.

4.2.2 The Simplest Neural Network

The simplest ANN is the perceptron and consists of three parts: the input layer, the hidden layer (or neuron), and the output layer. The input and output layers are intuitive to understand. They are analogous to independent and dependent variables in other models. The most difficult part of the ANN model to understand is the hidden layer. It takes the inputs and estimates the weights to multiply each of the variables to pass into the function, or "neuron." Critically, to tune these weights the ANN is trained on pre-labeled training data that is the basis for optimizing the relationship between inputs and outputs. The neuron in the hidden layer sums those input values, multiplied by their weights, and passes them into a function. The early perceptron used a step function called its "activation" function — adopting the neuron parlance of getting it to "fire." In Figure 7, a NN can be used to predict whether a building is a bungalow based on physical, parcel, and neighborhood characteristics, with the model being trained on existing data that structures the function. This functions very similar to a logistic regression model.

Figure 7. The three components of a neural network

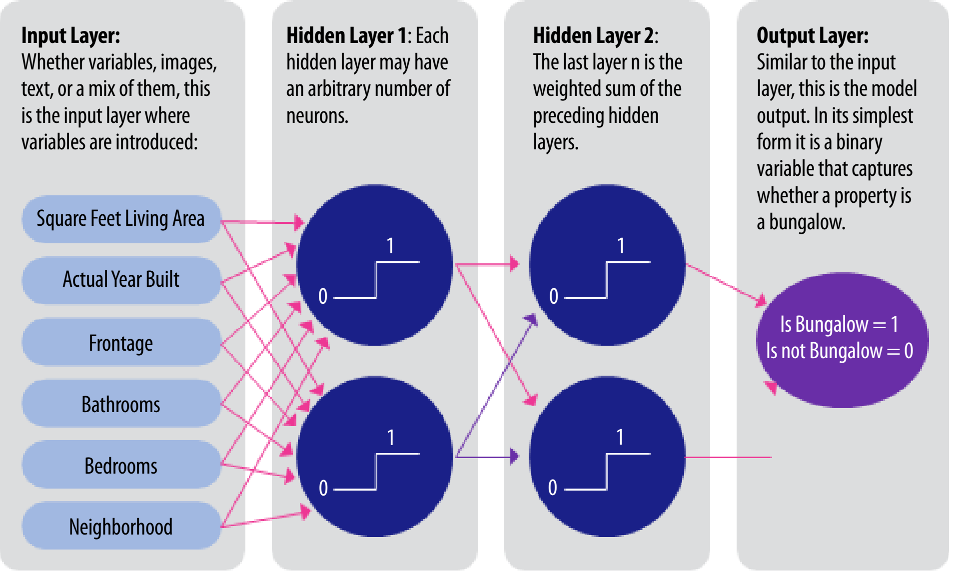


In addition, perceptrons may be adapted to predict continuous output estimates such as property values and perform similarly to regression functions. By substituting the identity activation function for the step function, we are able to estimate a continuous output. Some readers are likely to notice an interesting fact: What we have in substituting the identity activation is a general case of MRA that assessors are familiar with. The main difference is just in the way the "weights" or the "coefficients" and "intercepts" are estimated. Activation functions will be further discussed later. As mentioned earlier, ANNs are not limited to Boolean relationships. Many different functions can be used. In fact, they may be adapted to predict continuous estimates, such as property value, and can function like MRAs and logistic regression.

4.2.3 Wide and Deep: Using Multiple Neurons and Hidden Layers

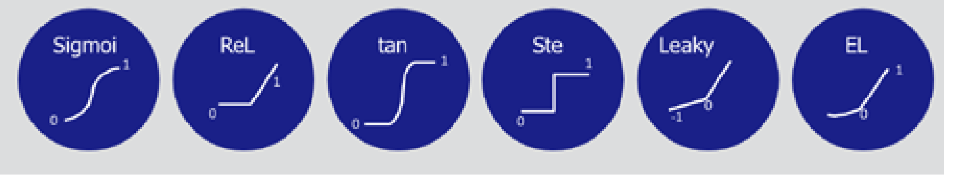
Our simple function above can be made more complicated and allows for more complex relationships between the input variables to emerge. It turns out that the early theory was correct: The ANN models are adaptable, whether there are two or 200 neurons or many more. More complicated relationships between the input variables can be created by "widening" the hidden layer through adding more neurons. In a comparatively simple two-neuron example (figure 8), one neuron may capture the interplay between, say, roof type and frontage to help distinguish a bungalow from other styles having a single story, such as ranches, ramblers, and Cape Cods. It is not a large leap to see how having additional neurons allows the model to learn complex relationships between variables.

Figure 8. Adding depth and width



The ANNs are flexible regarding the number of hidden layers, though there is a computational penalty, as discussed earlier. As it became computationally feasible to add more and more hidden layers, the concept of DL emerged.8 DL enables ever more complex relationships among the input variables to be formed, including ones that seem very human-like, such as image and speech recognition. Back to our example (e.g. predicting whether a residential property is a bungalow): DL allows for complex interactions between complicated variable relationships. In our example, the neuron in the first hidden layer effectively optimizes the year built and number of bedrooms together, and frontage type in the second neuron. The next hidden or deeper layer would allow the model to optimize combinations of those two inputs. The layers are finding the variable interactions and the interactions of interactions that best drive the prediction.

Figure 9. Some activation functions sometimes used in the hidden layers



The DL structure is powerful and changes modeling as it takes the work of creating custom variables out of the hands of modelers and puts it into the ANN model. The analyst becomes more of a higher-level manager. Instead of needing to predefine data transformations as David L. Jensen suggests in his "Tricks of the CAMA masters" series, the DL, sufficiently parameterized, finds those transformations on its own after being trained on test data.

Important in facilitating the development of these interactions are the activation functions, which act as gatekeepers for each neuron. Along with the step function, common activation functions are shown (figure 9). The "sigmoid" activation function is the form known to many modelers when used in logistic regression. The hyperbolic tangent activation function, frequently referred to as "tanh," falls in between the sigmoid and the "step" functions, steepening the transition from 0 to 1, but not as starkly as the step function. While not displayed here, the "softmax" function is a generalization of the sigmoid for multiple categories. These were important additions that allowed for nonlinear combinations of the inputs, or basically the ability of the models to identify "curves" in data.

The rectified linear unit, often referred to as "ReLU," and its relatives the "leaky ReLU" and the exponential linear unit "ELU," were important additions as models add hidden layers. ReLUs ensure that gradients do not get excessively large or small in magnitude. Essentially, they facilitate the optimization process while still allowing for rich interactions between variables.

Moving to more specialized formats of NNs, we can extract conceptual abstract and cognitively challenging ideas such as spatial or temporal dependence — with this we introduce the convolutional neural network (CNN) and its powerful image-processing abilities.

4.2.4 Convolutional Neural Networks

As seen in the earlier case study, CNNs are extremely useful to assessment and valuation organizations because of their ability to extract physical information from images and to classify images with higher-order structures. CNNs bring DL to the forefront of image classification and computer vision. Yann LeCun introduced the concept of convolutional networks in 1989 (LeCun, 1989) and with several collaborators released an implementation called LeNet-5 (LeCun et al., 1998), capable of reading bank checks. Nevertheless, it was Alex Krizhevsky, Ilya Sutskever, and Geoffrey

E. Hinton's AlexNet (Krizhevsky et al. ,2012) that brought widespread attention to the CNN. The AlexNet model demonstrated revolutionary performance gains over nonneural network image classification models with an 18 % reduction in error. While the architectures continue to get deeper, the conceptual means by which CNNs work remains the same.

How do CNNs work? Rather than taking each pixel of an image and passing it to one of the activation functions noted earlier, as would be the case in a multilayer perceptron (MLP) model, the CNN introduces two main innovations and is more efficient. The first is the convolutional layer, from which it gets its name, and the second is the pooling layer. The convolutional layer acts much like simple cells in the visual cortex, while the pooling layer acts like complex cells; their roles are, respectively, to activate based on simple patterns and to aggregate large patterns from simpler ones (Sejnowski, 2018, 131). The convolutional layer identifies smaller elements contributing to the whole, while the pooling layer identifies the essence of what they form cumulatively. The CNN is literally building the structure of the image out of the individual pixels and can identify components such as building walls, material type, or roof shape.

The convolutional layer works by taking several filters and sliding them over the input. The pooling layer works by sliding a moving window over the input and summarizing the value. The stylized structure of the model (figure 10), where an image is passed into the model, is then fed to a sequence of convolutional layers and pooling layers, which outputs the building grade estimate. Each successive convolutional layer applies a more complex filter, allowing it to extract more sophisticated features. The pooling layer associated with it then reduces the image size, leaving the most defining features and providing a smaller canvas across which the filter slides the next time. The filters are what pick out specific features. Filters from layers one, seven, 20, and 131 are shown below in figure 11.

The filter in the first layer has imperceptible patterning, while the filters from subsequent layers highlight other elements. The first convolutional layer starts to contrast the building from the lawn. The seventh convolutional layer starts to extract roofline features. The 20th layer starts to highlight window types. When the image is convolved and pooled more than 30 times it ceases to be recognizable as a house. At that point, it is identifying building features to determine grade.

That the layers build on each other becomes more apparent when adjacent layer filter activations are applied (Figure 12). Applying a filter from each of the sixth and seventh layers of the model to two images, it becomes apparent that the sixth layer filters are working to detect cornice lines. The seventh layer filters identify more detail from the roof and cornice lines and start to reveal elements of wall and siding type. Looking at the filters themselves does not immediately suggest that they would be able to identify these elements, but when they are applied consecutively to the images, they enable the model to see and act upon these objects

Figure 10. The CNN structure

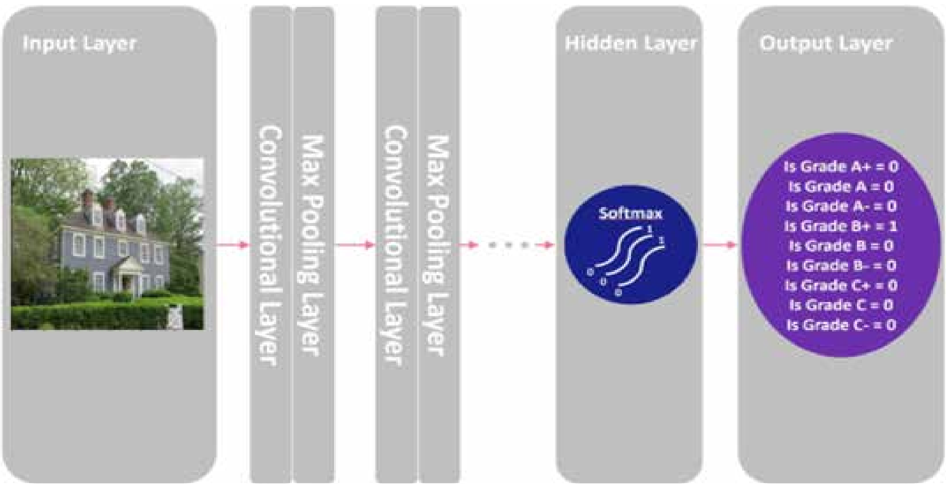


Figure 11. CNN filter output

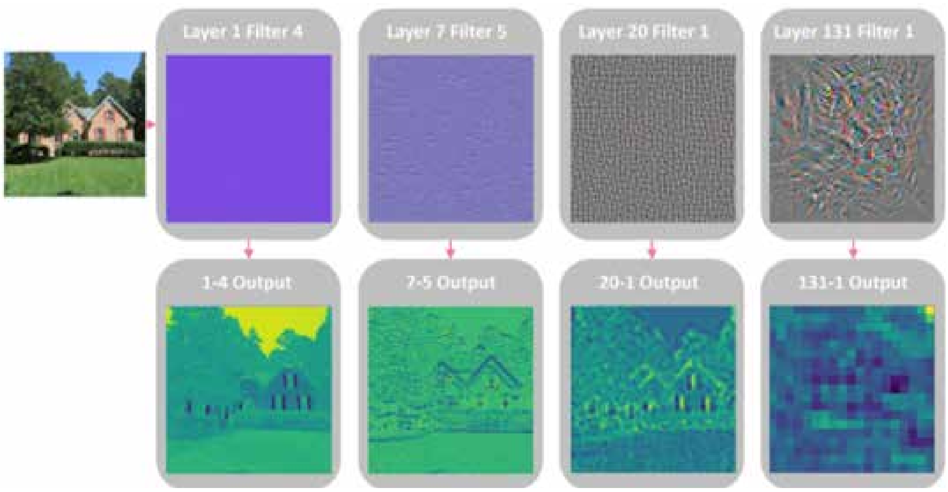
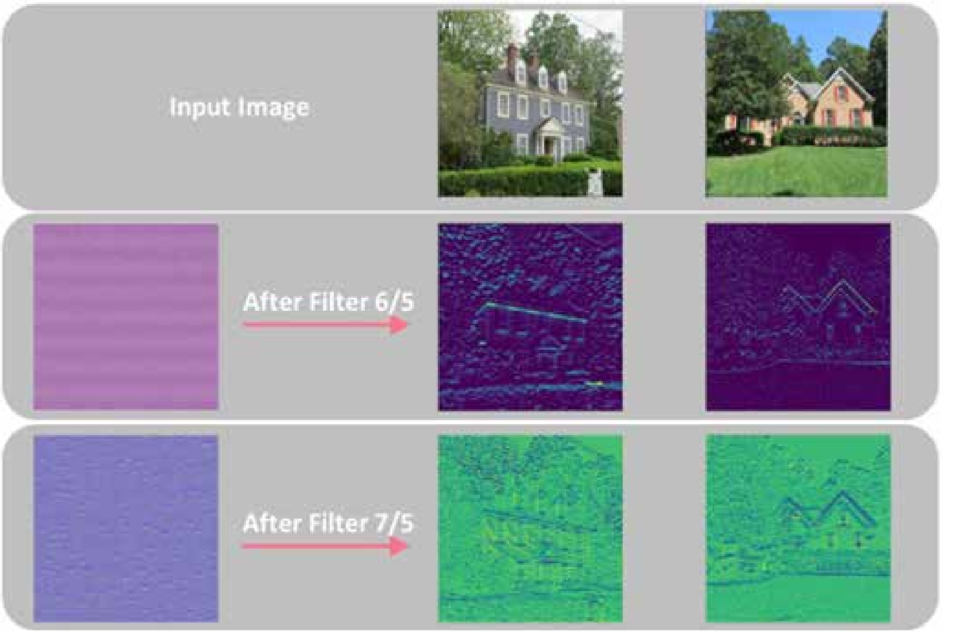


Figure 12. Filters and additional layers create higher-order structure



4.2.5 Recurrent Neural Networks

While image processing is seminal to CNNs, they can also be used for sequence-based analyses such as language processing, video, and time series. Recurrent neural networks (RNN) come in several flavors, and they can be used to generate an output estimate for a single point in time or for a time series. The primary difference is the dimensionality of the inputs. Instead of being a two-dimensional image (e.g. for CNNs), an input such as video is consisting of two-dimensional images in a sequence of image frames, and is processed in a RNN using a time-oriented convolutional layer placed deeper in additional layers. For time series, the input is a one-dimensional convolution layer while sequences make use of memory structures.

RNNs also have ability to flexibly handle different sorts of temporal cycles such as seasonality and catastrophic events such as tropical storms using a specialized RNN called short-term memory networks (LSTM) RNNs. LSTM is a class of gated RNNs, where the neural network has "gates" in that act like filters, letting information pass or drop. These gates let the model "remember" or "forget" things from earlier events in the sequence or time series. RNN carry some similarities with classic autoregressive models but is more flexible and handles multiple input variables.

4.2.6 Generative Learning

Generative adversarial learning (GAL) is an interesting and recent innovation using unsupervised learning to generate models similar in character to the original training data sets. For example, our training data set could be old maps and our output results would follow the structure of the inputs (e.g. modern GIS maps would be output with old map type colors, text, and visual structure). GALs consist of two parts: a generator and a discriminator.

There are several potential uses for GAL in assessing offices. Generators, by construction, can produce realistic outputs, even where the real output does not exist.

Trained generators can provide visual typologies and augmenting data, such as a map style or building style, for training staff in smaller jurisdictions. The generator could be used to produce a standard image for each possible combination of building grade, style, and other qualitative characteristic a jurisdiction wants to have in its field manual. Alternatively, where only a few examples of a particular style, siding type, roof type, or the like are available, the generator can "manufacture" examples for use in training other models.

The discriminator, on the other hand, is particularly good at identifying problems. As it is trained to recognize when the input is not a real instance of the concept, it can be used to identify misclassified or labeled data. As an example, if the improvement on a parcel is labeled a bungalow in the CAMA system, but the image and characteristics suggest a ranch, a trained discriminator could be used to detect the mismatch.

4.2.7 Conclusion

With the assertion that DL can replace or complement the important contributions expert modelers provide, it is reasonable to ask why deep learning models have not taken over. The short answer is that setting up DL models to achieve an expected outcome is difficult and requires a lot of data, and the results are uncertain and more difficult to interpret. It is one reason that boosted methods still dominate in areas such as estimating parcel value, where they can approximate a fair amount of Jensen's prescriptions (2013, 2015) without a great deal of fanfare. In spite of these limitations, DL models are rapidly gaining traction in the industry and hold great promise in assessment and valuation workflows because of their data and modeling flexibility, and their potential simplicity compared with other ML methods.

4.3 AI Search and Optimization

Most AI computing algorithms, including ML, were developed in the field of computer science for efficient execution on modern computers. The prior sections on ML introduced several numerical computing techniques where the computer learned an NN or tree representation through ML training procedures known as search and optimization. AI search and optimization procedures evaluate many possible numerical solutions quickly to find a solution, or a set of numerical parameter values, that satisfies some formula or constraints. A simple example of a numerical search problem is solving for a system of equations where you seek to find values for the set of numerical parameters that make each equation in the system true simultaneously.

This section introduces the application areas where search and optimization techniques can create business value. ML has limitations in problems with complex symbolic business constraints. Search, constraint programming, and optimization techniques excel at removing limitations and complexity to deliver end-to-end automation, as with AutoML discussed in an earlier case study. Hearing scheduling, inspection planning and routing, comparable sales selection, and model calibration are four primary areas where AI search and optimization can be applied to improve mass appraisal decision processes.

AI search and optimization are useful for complex business tasks that need to examine many possible solutions and select the best alternative in end-to-end automation scenarios. AI search and optimization are used to calibrate automated valuation parameters and by AutoML techniques that select the appropriate ML architecture and associated learning procedures and parameters — a critical technology for widespread adoption of these techniques. Software that includes AI search and optimization procedures allows nonexperts to easily create robust predictive models. Such software also saves experts valuable time because the computer does the laborious trial and error scans through candidate model types and parameter choices that would otherwise make these unapproachable to all but mathematicians and computer scientists. Based on this section, ML, search, and optimization are really "three sides of the same coin," and search is one of the foundational legs in our discussion of AI.

4.3.1 Introduction to Search

Much of the early work in AI was in the area of symbolic knowledge representation (Sowa, 1983; Breuker and Van de Velde, 1994). Symbolic knowledge representation, such as a business rule in the form of if–else, is a way to specify or declare the problem and the desired characteristics of the solution. The key feature of symbolic computing and search is that the analyst, who is describing the problem and solution quality and goals, does not specify how to find the solution. The search and optimization methods find the answers to specific and complex problems without an explicit algorithm being programmed, giving search and optimization their places within the AI toolbox. In short, search and optimization allow the user to state the "what" that defines the problem and then, through search and other techniques, AI self-determines "how" to solve the stated problem for the user.

4.3.2 Search as 'Inference' in Rule Systems

One of the simplest applications of search is in the evaluation of business rules found in rule systems (Ullman, 1988). Users declare specific knowledge in symbolic form (e.g. symbols, sets, relations, formulas, programs) as the business rules and then leave the execution of those rules to the computer system. Rule systems perform "logical inferences" that are implemented internally as search procedures. Search procedures constitute the control design of the "inference engine" within rule systems within rule systems; as searches are executed, new facts are derived and queries are answered The controlled inference plays out in a highly iterative manner, scanning through the rules and over the state of the database (rule-base) repeatedly.

Knowledge representation for business problems is performed by users by specifying very simple conditional statements (e.g. if-else expressions). In other words, a problem is defined by logic syntax and semantics. Reasoning is performed by the computer by presenting data or facts about the world through an inference engine. Rule evaluation cycles are repeated over and over, with newly generated facts triggering new if-then fact creation, i.e., logical inference. We call this "forward inference" or "forward chaining." You may find this style of a forward chaining "if-else" rule system within your CAMA solution today. Typically, a rule base is used within workflow systems. Business rules interacting with the CAMA database are a common way to define workflows within a process.

There is an alternative and perhaps more powerful model of rule evaluation called "backward chaining." Like forward chaining, this model uses a search procedure for its inference engine. In this case, the system starts with the goal presented by the user. However, rather than letting data changes cause rules to fire and create new facts, the backward chaining reasoning engine acts more like a query that you may be familiar with from querying your CAMA relational database.

With backward chaining, users pose a logical assertion to the rule base by asking the system if a given fact or logical statement is true. The rule inference engine performs the reverse inference procedure. It starts by reasoning backward from the goal to find a solution (Jeffrey, 1981). With the goal stated, the search procedure seeks out "what needs to be proved or established as fact in order to prove the goal the user has stated." If the search procedure can find at least one path through the rules and the data (or facts) that establishes the goal, then the search procedure is said to have found a solution.

4.3.3 Symbolic vs. Numerical Search and Optimization Techniques

When looking at how these fields have developed, we can see significant overlap between the early work in numerical search and symbolic search in computer science. The overlap and interplay continue in this modern day and in the age of AI. The easiest way to frame the difference between symbolic and numerical procedures is to think about the structure of the problems being represented. Symbolic search and optimization methods are analogous to solving complex word problems in high school mathematics, whereas numerical procedures use math functions and expression only to declare the problem. Intuitively, we recall that mathematical word problems were more complex, since they required representation of relationships and other concepts between the problem actors and problem constraints expressed in English. Word problems (symbolic) require both logic and math, but many math problems (numerical) do not require logic or inference.

In the broadest sense, both techniques seek to find a solution, most often expressed as variable values or parameters that meet certain user declarative constraints and, at the same time, optimize the user-expressed quality attributes of the solution. The quality attributes are typically expressed as some form of score or objective function over those variables. Numerical optimization procedures typically search numerical parameters and, in some advanced cases, can search through spaces or forms of quantitative functions. Symbolic optimization procedures, although they can work with numerical domains, are more rooted in working with sets, relations, and symbols drawn from the specific problem as well as with symbolic expression syntax.

4.3.4 General Search and Optimization

The following sections will provide an overview of some of the more popular numerical and symbolic search procedures (Bolc and Cytowski, 1992) that can be applied to a wide variety of problems. They are important because they can be used for any type of problem and can leverage computational cloud computing platforms to speed up the running time to find good solutions.

4.3.4.1 Hill Climbing

We may often hear the expression "the curse of dimensionality" when discussing optimization problems. This refers to the fact that as the dimensions of the solution or problem are increased, i.e., the number of variables in the solution description increases or the size of those variable domains grows, the resulting size of the search space increases exponentially.

Hill climbing is a classic technique to avoid wandering around searching for a solution without enough focus or aim. Hill climbing is also called a greedy search procedure because in each step, it will take the path that improves the objective function the most. It stops when no more improvements can be made.

Nearly every ML procedure is solving an optimization problem during the model training phase. During training, ML procedures seek to minimize the error against the training examples. The gradient descent procedure is an application of a greedy search procedure.

Hill climbing and thus most ML training procedures are subject to the risk of finding only a locally good solution. In practice, the search procedure can converge to or become stuck in a local maxima or minima. Greed may be good, but it is not always the best.

4.3.4.2 Case-Based Reasoning

One approach to avoid the limitations of hill climbing and to solve certain problem structures is to use case-based reasoning. The basic idea with case-based reasoning is to keep a historical database of previous problems and their solutions. When a new problem is presented to a casebased reasoning engine, the historical case base is referenced, and the search procedure tries to match the new problem to a historical problem. That allows the search procedure to start potentially closer to the solution than if the procedure started blindly searching. Sometimes that is called a warm start or primer.

Once the most similar problem from the case base is identified, the search procedure then begins by using the prior solution as the starting point for finding the solution to the new problem.

Often very domain-specific techniques are used to create a new solution from the matching historical solution(s). Case-based reasoning helps avoid the problem of local minima or maxima.

Our industry already applies the case-based technique comprehensively. Consider the comparable sales approach to value: The most similar sales for a subject property are selected from a "case base" of historical sales, and then a "heuristic" valuation function is applied to the best match cases to estimate a value for the subject. Nearly all custom implementations of comparable sales selection programs found in CAMA systems today can be replaced with a standard case-based technology engine.

4.3.4.3 Simulated Annealing

Simulated annealing is an adaptation or essentially a heuristic approach that is embedded in standard search techniques. It is special in that its heuristic is a natural one and is motivated by analogies from physics in how materials physically anneal related to heat and heat treatment.

When material is heated and then slowly cooled, it forms a uniform structure. The idea of simulated annealing is to mimic this process with search and ML procedures.

The key advantage of simulated annealing is that, unlike strict hill climbing heuristics, simulated annealing allows for exploration in directions that yield a worse score against the objective function during the search process and thus opens the possibility that search procedures will not get stuck within a local minima or maxima.

Essentially, in simulated annealing search a probability function, sometimes called the energy function, is used by the search procedure to accept or not accept a next step. By introducing the energy function into the computer search process, search algorithms can avoid missing better solutions that may have been bypassed because of the limits of hill climbing search and its tendency to converge on local minima or maxima.

Simulated annealing implementations often require some control parameters, such as the probability schedule as well as the acceptance threshold at each probability stage. Like much of ML, these input parameters are procedure hyperparameters that also can be meta-searched and fine-tuned during domain-specific problem application.

4.3.5 Constraint Technology

The above search techniques are general in that they can be applied to essentially any problem structure. This section provides an overview of two highly industrialized applications of AI search that have been engineered to efficiently solve specific forms of problem structures. These specific applications are important because they often are accompanied by specialized analytic engines that are off-the-shelf software libraries directly embedded into business software applications such as the Google OR Tools library (Perron and Furnon, 2019).

These specialized search engines directly compute solutions from declarative specifications of the "assembly" problem. Once an analyst masters the problem specification process and format, they can be applied across many problem structures without the need to custom program the solution. In the abstract, the solutions to these problems do not need to be tailored to the specific industry if the knowledge can be modeled for the standard solvers.

4.3.5.1 Mixed Integer Linear Programming

Linear programming (LP) is an early optimization technique developed in the late 1940s. This technique optimizes a linear equation that is subject to linear constraints. A mixed integer linear program (MILP) is a form of linear programming in which some of the variables are constrained to be integers only. LPs and MILPs are often solved using a class of search procedures called "branch and bound" methods. These methods search through a candidate set of solutions by forming a search tree with the full set of possible solutions (represented by the root of the tree) and then avoiding fruitless explorations.

LP and MILP are much more powerful than standard MRA in that they can express any form of linear constraint over the solution variables, and they are not limited to expressing only a specific individual variable bound, which is the state of practice in today's CAMA systems. This allows LP and MILP solver engines to consider the proper balancing and tradeoffs in a much more automatic fashion versus today's traditional two-stage interactive approach to model calibration and post calibration analysis, which is more laborious, time-consuming, and error-prone. MILPs offer these benefits without losing the understandability or explainability that occurs with parameter-less models calibrated with ML techniques.

4.3.6 Constraint-Satisfaction Programs

With LPs and MILPs, the search procedures apply only to problems whose decision variables are numbers; they cannot be used to solve problems that include symbolic constraints. Symbolic constraints are extremely useful for more complex search problems found in more general business and operational decision-making. Typically, constraint-satisfaction programs (CSPs) are used to find good solutions (decision variable values) that are subject to many complex business constraints.

Three classic problems in the appraisal industry can be targeted and solved by CSPs: 1) hearing scheduling, 2) inspections scheduling and routing, and 3) comparable property selection.

To understand the declarative structure of the constraint problem and the decision variables, consider a complex hearing schedule. This scheduling problem has inputs such as a set of protests, time needed to adjudicate case types, property attributes like building class, and the desired or required length of the session. The solution to the hearing scheduling problem, called the decision variables, provides a hearing venue closest to the taxpayer's property or other location, and a schedule that honors all the business that needs to be accomplished. Although it is possible to make procedural computer programs to analyze these tasks, those programs are generally brittle and do not have the ability to generate multiple solutions easily.

The advantages of CSPs are many, but the primary one is the ability to deal with symbolic constraints in language form, such as "these assignments must all be different" and “at least five of the values must be drawn from this set,” or any other complex combinatorial properties. Unlike LPs and MILPs, which are rooted in numerical parameters and numeric relationships, CSPs are rooted in the set theory and relational theory that underlie relational databases.

4.3.7 Evolutionary Search Methods

Of the search methods presented previously, only simulated annealing used randomness in the search procedure. There are wide bodies of knowledge and application that focus on evolutionary search and optimization procedures. Like simulated annealing in intent, the family of evolutionary search methods and optimization procedures have the common theme of harnessing randomness to avoid solutions that converge to a local maxima or minima. Like the methods above, evolutionary methods span both numerical and symbolic problem structures.

4.3.7.1 Evolutionary and Genetic Approaches to Search

Originally, evolutionary search procedures were motivated from the optimization concepts found in nature and Charles Darwin’s theory of evolution. The core ideas of biological “procreation,” “inherited variability,” and “survival of the fittest” are leveraged within the design of this class of search and optimization procedures.

In the 1960s and 1970s, German researchers pursued computerized methods for optimization and developed search procedures that leveraged randomization (Schwefel, 1965, 1995). The key innovation behind early approaches to evolutionary search was the addition of a process in which good candidate solutions were combined in some manner to generate even better solutions. In the mid-1970s, these evolutionary methods were extended by American researchers into what is today commonly referred to as genetic algorithms (Holland, 1975, 1992). These methods hold promise to produce outputs that are more easily interpreted and understood by business people and domain experts rather than just ML or statistics experts.

5. CHALLENGES: INTEGRATING AI WITH CAMA

This section discusses the challenges and state of the art of applying AI to the CAMA software that is so critical to the modern assessment and appraisal office. This section surveys several product features designed to introduce ML and constraint optimization tools to traditional users of CAMA systems. It outlines selected AI product enhancements made to a commercially available CAMA system from Tyler Technologies (iasWorld CAMA), including the context and motivation for these changes. It also addresses factors blocking industry acceptance and adoption of AI for real estate valuation, despite a well-established body of research, and discusses recommended methods, including gradient boosting.

5.1 Valuation Technology and AI in Mass Appraisal Systems

This section discusses a design philosophy for improving trust, transparency, usability, and simplicity when implementing a CAMA system with AI — critical for timely adoption by public administration organizations. A vendor-oriented case study is presented in section 5.5.

Over the last 20 years, the core techniques used by valuation practitioners have not changed much, despite research proposing the application of ML, geospatial statistics, and other advanced procedures that improve accuracy. Approaches to valuation in commercial systems today were developed during the 1970s through the mid-1990s, and the industry’s approach to its statistical valuation methodologies has remained remarkably stable (Dasso, 1973) except for a few dynamic areas such as GIS-based improvements. This stability is in direct contrast to private sector valuation, where practitioners such as the mortgage industry use a broader and more diverse array of mathematical and AI approaches to valuations.

Local government valuation offices did not implement AI methods despite their decades of experience with data quality control, statistical analysis, and model development. It took the recent advance of open-source software, cloud computing, and the new digital ecosystem to reduce the cost and complexity of initiating AI applications.

5.2 AI Research in Mass Appraisal

Results achieved in the research literature generally have established AI's superiority over classic MRA modeling. Most studies primarily measured and ranked AI methods in terms of error metrics. Notable improvements in model error have generally been achieved with the straightforward application of common AI techniques such as supervised forest learning and multilayered neural networks to real-world data attributes and markets (Zurada, 2015; Lowrance 2015; Antipov and

Pokryshevskaya, 2010). Unfortunately, the improvements alone are not compelling enough to drive adoption except in a few leading assessment organizations. AI research has so far not addressed the critical issues of interpretability and defensibility.

5.3 Challenges and Barriers to Adoption of AI

A likely barrier to AI adoption is the comfort and utility of the classic MRA forms, which are effective and offer intuitive formats. These transparent approaches also benefit interpretability and value defense and produce property valuations that are both fair and equitable. The fair and equitable taxation requirement places preference on models that might be less accurate overall but that produce errors that are equally distributed and not overtly progressive or regressive on critical attributes. In addition, calibrating AI models is a challenge, and models that merely optimize predictive accuracy at the expense of other important considerations must be avoided. Aside from traditional MRA and sales analysis, there is very little AI valuation research that outlines how to modify AI learning parameters or routines to improve model equity.

5.4 AI Technology Gaps in Mass Appraisal Systems

AI research has not been widely incorporated in commercial CAMA solutions. This is probably because of the barriers to adoption and the potential complexity of AI model calibration over MRA methods. Furthermore, three research and methodology gaps add risk and complexity:

- How to control and adjust training parameters of AI models to achieve stability, avoid bias, and ensure equity
- How to adjust training and data regimes to incrementally improve sales ratio metric performance and year-over-year benchmarks.
- How to adapt and present AI techniques in appeal defenses.

5.5 Product Design Approach and Philosophy for CAMA Software

In section 5.5, the reader will learn more from Joe Wehrli at Tyler Technologies about product design and philosophy issues for handling AI integration in their CAMA products.

Our product design approach with respect to AI has been governed by a few key tenets at Tyler Technologies. The first is to gain user trust and understanding without departing too far from existing modeling paradigms. The second is that workflows cannot be more complex than the existing processes used to calibrate MRA models. Third, the product design should mitigate or even remove the risk introduced by application of opaque, nonparametric models and favor reasonable economic interpretability when settling property tax disputes and reviewing appraisal practices.

5.5.1 MRA with AI-Based Feature Selection

The core of mass appraisal market modeling for the last several decades has been a constrained multiple regression approach. These models are calibrated primarily from real estate sales information and, in some cases, combined with geostatistical techniques to calibrate location-aware regression models. The system includes functionality for model variable definition and transformation. It also supports automated feature selection using standard stepwise techniques. The process still requires strong skills by the modeler as well as trial-and-error efforts to settle on an appropriate feature set.

With a review of modern AI techniques, it quickly became apparent to us that traditional feature selection is no longer required in the classic sense and AI can simplify use cases. Several methods might assist in simplifying MRA feature selection. One arrived at through experimentation was the standard gradient boosting method (Friedman, 1999), and another was some advanced formulations of MILP (Bertsimas, 2018).

Since our goal is to build trust in new approaches, a potentially effective way to do that is by embedding new AI-based techniques in the blanket of traditional methods. This affords users low-risk opportunities to use, understand, and grow to trust new methods. In this case, we developed a proprietary thresholding algorithm that uses gradient boosting for variable selection. The approach ranks and determines a cut point used to exclude variables from the MRA calibration procedure. Variables that don't materially improve the gradient boosting models are candidates for model exclusion to improve MRA model efficiency.

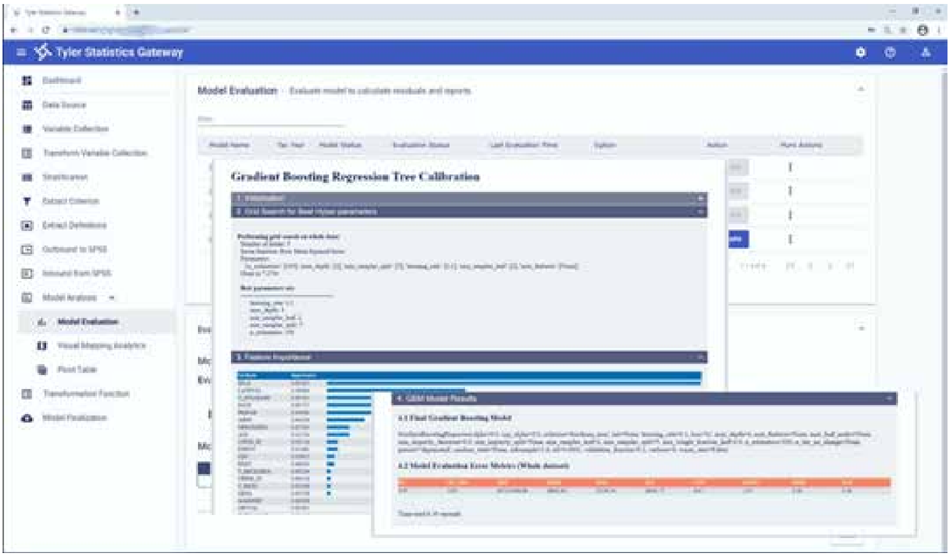
The approach is extremely effective and anecdotally performs better than stepwise procedures. Also, stepwise procedures that rely on correlation metric-based heuristics tend to suffer in the presence of collinearity and other violations to the assumptions of multiple regression. Using gradient boosting (GB) for variable selection sidesteps many of the problems that adversely impact classic subset feature selection and is much faster and more automated.

Finally, this feature design produced by simple GB often has better predictive performance than the target MRA model being worked by the user. Since gradient boosting is not limited to either smooth or linear fitting, its error metrics often serve as a "teasing indicator" of the less flexible MRA model being calibrated. The idea of this approach is to position and motivate users into trying AI, i.e., GB, as the primary method for model calibration discussed in the next section.

5.5.2 Gradient Boosting Method

Based on an internal review of existing research literature and additional proprietary factors, we determined that implementing regression tree learning using GB as our initial AI-based, nonlinear, nonparametric model calibration feature would provide users a straightforward and low-risk path to break free from the mass appraisal industry's linear and parametric optimization shackles (figure 13)

Figure 13. GB model training performed within market modeling workflow



GB as an approach to property valuation model calibration offers several tangible benefits to government appraisal users. One of the main business benefits is that GB is more robust against statistical assumptions and problems that challenge our traditional MRA analysts, such as multicollinearity, outliers, data errors, and progressivity/regressively. This translates to potentially simplified feature engineering processes. GB affords the opportunity to greatly enlarge the market segmentation used in building real estate valuation models over what has been traditionally done with MRA. Analysts will need to adapt to take full advantage of GB. Different skills will be useful to help with interpretation and error analysis and to impact model evaluation metrics that are not directly incorporated into the GB learning rules.

Although the GB method appears to produce good models comparable with MRA, and it has potential to greatly simplify the modeling workflow, a product design still requires additional features to assist with validation and interpretation of the opaque models that ML produces.

5.5.3 AI Validation and Interpretation

Several possible feature designs, many still experimental, are discussed in this section. One of the primary barriers to broadly adopting AI is the lack of an interpretable model form. We propose a multifaceted design strategy that begins to address both validation and understanding of "black box" models that for the most part are uninterpretable. It is a hybrid of existing techniques that assessors are comfortable with, mixed with new, AI-based features.

Other AI-related techniques such as mixed integer linear solvers and genetic search engines (Wehrli, 2019) can be used as model surrogates for smaller stratifications in parallel with GB to help validate that the GB models do not have undesirable properties, such as poor generalization caused by overtraining or bias, and jaggedness that can result from problematic distributions of training data.

IAAO has created standard ratio review methodology that prescribes a set of metrics to analyze and interpret model error, sales ratios, and vertical and horizontal equity. These tools can be applied independently of the model learning method, so we built a new interactive pivot table tool that allows users to interactively segment and partition any of the standard ratio analysis metrics from MRA models, GBM models, and other emerging types of calibrated valuation models.

5.5.4 Conclusion

This section has briefly summarized the challenges and barriers of leveraging AI within commercial mass appraisal systems. Factors impacting industry progress are both cultural and technical. Conservative product design is used to address these barriers and challenges. We describe enhancing our existing commercial mass appraisal system using a combination of traditional and new AI methods. This product design builds user trust without forcing them to throw out existing knowledge and processes. A hybrid approach may allow users to transition to AI modeling as they gain trust, experience tighter GB model fits, and reduce the model counts needed within a jurisdiction. More work is needed to simplify the modeling process by providing easier interpretations.

6. RECOMMENDATIONS

This white paper reviews methods and case studies for the application of ML and AI to tax assessment. It offers insights on the current state of the art and some of the key concepts, processes, and technological developments that jurisdictions should consider for improving their property taxation systems. The goal for the integration and adoption of AI and the related technological ecosystem is to enhance prediction accuracy, reduce costs and time, and prove trustworthy. The adoption of AI within CAMA-based systems needs to be explainable and defensible.

The role of AI and ML in the property field is not new. CAMA-based systems and automated valuation techniques have continued to evolve. The development of geostatistical and ML approaches and the continued evolution of data digitalization and workflows drive the changing nature of automation within the industry. Current IAAO standards on mass appraisal concentrate on statistical algorithms and equation-based applications. In terms of AI adoption, AI and ML-based research has strongly positioned "explainable" algorithms to be used moving forward.

Adoption for most assessment jurisdictions should focus on AI methods that are transparent to the modeler and explainable to the taxpayer because of the lack of tangible equations in AI. These explainable subsets of AI methods are powerful; however, they tend to be focused on driving efficiencies, and this reduces the scope of potential competencies of AI. The adoption of AI methods is not yet a standardized process, which means each jurisdiction or assessment office will need to review its current level of technological adoption and adaptation and evaluate how to incorporate these technological and analytic trends moving forward.

Assessment organizations have many potential business processes and problems that will likely improve with AI implementation, and we included a list in Table 3 for your reference as you consider pilot projects and areas of innovation.

Finally, the case studies on Property Valuation Services Corporation (PVSC) of Nova Scotia, Canada, BC Assessment, Rwanda, Nigeria, Uganda, and Zambia demonstrated success in applying AI for valuation and assessment purposes. Their accomplishments illustrate that the path to AI adoption requires a strategic and decisive process of implementing the correct technology and developing the necessary human resources and varied skill sets. Additional recommendations emerging from the report are discussed below.

Table 3. Possible AI Pilot Projects

Business Improvements	Areas	Possible AI Methods
Open, transparent, intelligent services to clients/ citizens	• Automation of constituent profile creation • Constituent services (e.g. chat bots and email requests) • AI agent-based e-hearings and e-informal reviews	• AI knowledge graph/search engine • Chat bot/natural language engine • Caseand rule-based decision engines
Improved quality of data and data management processes	• Flexible reporting and outlier analysis • Auto flag discrepancies between CAMA records and imagery, and GIS • Flag discrepancies in CAMA records for autoassisted outlier review and change review • Speech-based field data entry	• Caseand rule-based decision engines • Convolution networks / deep learning • Natural language engine • Speech recognition engine
Improved quality and flexibility of workflow processes	• Process management and control of decision problem automation within workflow • Automation of permit ingestion, transfer ingestion, exemption processing, and fraud detection	• Robotic process automation • ML models • Optical character recognition • Natural language engine
Improved resource planning and optimization	• General automation of business planning and scheduling (e.g. tax-cycle optimization) • Optimization of goal-based labor/office resource planning and budgeting • Data-audit planning • Field-inspection and desktop-review planning	• Constraint satisfaction programs • Mixed-integer linear programming
Reduction and removal of labor costs for data collection and maintenance Shift to mass data collection driven valuation	• Auto-capture physical information from imagery and GIS data (e.g. building grade and condition) • Auto-capture and ingress thirdparty data and imagery services • Creation of 2D and 3D building sketches • Selection and inference of contextual GIS features (e.g. corner lot and waterfront) • Interpretation of MLS listing narratives and internal imagery	• Convolution networks /deep learning/ recurrent neural networks/ Natural language engine • ML models

6.1 AI Techniques, Data, and Digitization

- Assessment organizations need to adopt new technologies to increase efficiency and embrace new value methods. The digitization of data requires operational attention, as adopting these technologies and analytic methods is essential to improving and enhancing tax administration and appraisal moving forward.
- Jurisdictions and assessment offices must identify the role of AI in appraisal and assessment for business problems. This relates primarily to data and quantitative methods that improve appraisal accuracy and data capture techniques. Whether the AI methods increase accuracy, transparency, and efficiency are issues that jurisdictions must address prior to AI adoption.
- Jurisdictions should recognize that AI applications extend beyond appraisal and assessment. AI provides the opportunity to improve operations, workflows, and data quality as part of digitalization.

- AI is based on pattern and image recognition is well-suited for automating the mass collection of physical attributes and cadastral features of property. This offers the opportunity to value significant physical features (such as building attributes) within ML-based training models. Remote sensing technologies are becoming less expensive for capturing parcel and building information and there are now applications and technologies that combine geospatial data and analytic methods such as change detection using imagery that can augment appraisal accuracy. Thus, the application of digital and remotely sensed information incorporating image-based attributes can potentially improve predictive accuracy and reduce the cost of establishing and updating tax registers by allowing automated generation of building footprints and cadastral information.
- One of the challenges with AI integration is the role of data and the related governance and quality. To adopt AI for improving operational capacity, there must be enough appropriate data available to train ML applications because AI is data intensive. Even in instances where data is available, it may not be accessible or organized readily for AI applications or may need extensive tagging and review. The integration of AI requires an enabling environment, one that relates to the incorporation of AI "systems" and organizational values. This requires a coordinated approach for the wholesale adoption of AI and ML within public administration offices, including staff training and education.

6.2 Implementing AI in the Workplace to Improve Productivity

- The successful adoption of AI necessitates a good management structure to deal with the uncertainties of the technology's applications and to fully engage with the integration. To adopt and modulate AI is initially time-consuming and may require high entry costs. The integration of your technology stack with AI capabilities will be important.
- To fully embrace AI, it is necessary to build a diverse talent pool and to align the correct people with the correct skill sets. Traditional job classifications may evolve. Staff and vendors must work together to implement AI technology.
- The integration of technology also requires that our profession create additional standards for effective AI use and efficient working practices. This requires a clear strategy for everything from designating protocols for scaling up data analysis, to establishing and consolidating platforms and the analytic ecosystem, to codifying standards for data lineage and veracity.
- Because technologies are advancing rapidly, the successful scale-up of AI requires the development of an AI "ecosystem." Within the appraisal and assessment community, valuers, data quality specialists, ML engineers, data and analytic officers, and data modelers will need to help foster the successful application of AI and address gaps in the skills required for fruitful implementation and delivery.
- There is a practical difficulty in transitioning to AI: the lack of a blueprint or standard AI road map. Incorporating AI is different from traditional software implementation, thus changing the status quo requires agility, openness, resources, planning, and trust. This may initially seem disruptive; however, the power of AI is in complementing and augmenting human capabilities, not replacing them.
- There are many proven, low-cost AI options to buy off the shelf. It is therefore key to take stock of what already exists within the organization's value chain and customize new technology as needed to provide the basis for successful scaling-up.

6.3 Education and Training

- Education is an essential ingredient for AI uptake and scalability. The staff needs to understand the organizational role and outputs from these models. Active collaboration between local professional bodies and international organizations to undertake member training and development is fundamental to increasing the usage of AI techniques and AI-based technology.
- Clear qualifications and standards for practitioners will create confidence in AI technology. Organizations like IAAO will play a role in standardizing employee education and knowledge. It will be important to embed education and training requirements to increase AI literacy across organizations. This is an essential pillar for building confidence in AI and driving adoption and usage. Building AI literacy needs to be a cross-functional focus.

6.4 Ethical AI: Ensuring Privacy, Standards, and Transparency

- The growing awareness and consideration of ethical challenges from biased training data requires the development of a data governance and AI framework to help regulate AI-driven technologies and implementation. It is imperative to establish how AI systems arrives at a given outcome and take decisions out of the black box. A clear governance framework and ethics committee can help develop practices and protocols that ensure their code of ethics is properly translated into the development of AI solutions.
- AI affords tremendous opportunities, from increasing efficiencies to improving predictive outcomes. Against this backdrop, it will have impacts upon equity and fairness, and this raises important questions around ethics, trust, legality, and responsibility. This necessitates the implementation of specific technical guidelines to ensure that AI systems are safe, transparent, and accountable.
- Engagement with AI, and particularly engaging people with AI, will increase acceptance and trust in ML. Organizations, particularly assessment offices, must design their AI strategies with trust in mind. This necessitates building the right governance structures and making sure ethical principles are translated into the development of algorithms and software that are publicly accountable, transparent, and explainable.
- There are several regulatory problems with AI. For instance, research and development activities require little physical infrastructure for the various components of an AI system. And AI programs and software may have data or software components taken from multiple libraries, each of which is built and developed discretely. Few, if any, AI systems are built from the ground up: They do not typically use components and code that are wholly the creation of the AI developers. Modern computer systems also use commercial off-the-shelf hardware and software components, most of which are proprietary.
- The inner workings of and the interactions between the components of an AI system, akin to expert rules-based systems, can be as opaque as those of earlier technologies. Critical features underlying an AI system's operation thus may not be immediately apparent or readily susceptible to reverse engineering. As described earlier, the use of complementary and comparative ML techniques and statistical studies helps calibrate and validate output accuracy.
- AI's opacity is its most challenging aspect, as users may not be able to detect potential issues/features of an AI system, and they may not be susceptible to

reverse engineering. This autonomous nature of AI creates issues of foreseeability and control. Defects in the design of a complex AI system might be undetectable and taken together these characteristics present difficulties for accuracy and scalability. Data quality control, data sampling, and assessment of outputs are required for validation and reliability.

6.5 Adoption and Uptake

- While we believe data integration, automating processes, and use of technology have great potential, there are few examples where these processes have been successfully implemented for mass appraisal exercises. Therefore, the adoption of AI needs to be undertaken within a compatible and accessible infrastructure.
- Each assessment office or jurisdiction must establish where it is on the "AI curve." To commence the scalable AI process, jurisdictions must strategize starting with the premise that "the front end is the data, and the back end is the defensibility." Improvement opportunities should be prioritized and sequenced based on value and change requirements, along with an organization's capacity for implementation and change.
- The adoption of AI (in the first instance) requires collaboration on the technical and software aspects of implementation for comparables and valuations.
- While the use of AI is not widespread, some jurisdictions have adopted its use in mass appraisal and mass data collection. Jurisdictions can learn from the experience of those who have implemented or have tested the use of AI.
- Finally, widespread AI adoption and uptake requires a value system in your organization and communities predicated on the following qualities: Predictability, Transparency, Understanding, Control, Fairness, Equity, Morality, and Trust.

7. REFERENCES

Ali, D.A.; Deininger, K. and Wild, M. 2018. Using satellite imagery to revolutionize creation of tax maps and local revenue collection. Policy Working Paper No.8437. World Bank, Washington, DC.

Antipov, E. and Pokryshevskaya, E. 2012. "Mass Appraisal of Residential Apartments: An Application of Random Forest for Valuation and a CART-Based Approach for Model Diagnostics." Expert Systems with Applications. 39(2): 1772-1778.

Awrangjeb, M. et al. 2018. "An Effective Data-Driven Method for 3-D Building Roof Reconstruction and Robust Change Detection." Remote Sensing. 10(10): 1512. Found at: https://www.mdpi.com/journal/remotesensing/special_issues/Building_Detection.

Balogun, T. F. 2019. "An Assessment of Property Tax Administration in Edo State, Nigeria." The Indonesian Journal of Geography, 51(1): 69-77.

Bengio, Y. et al. 2006. "Greedy Layer-Wise Training of Deep Networks." In Advances in Neural Information Processing Systems, eds. B. Scholkpf, J. Platt and T. Hoffman. Curran Associates.

Bertsimas, D. 2018. "Machine Learning Under a Modern Optimization Lens." http://www.lnmb.nl/conferences/2018/programlnmbconference/Bertsimas-1.pdf.

Blake, M., and Kriticos, S. 2019. Enhancing property tax compliance in Mandalay. International Growth Centre. Found at: http://eprints.lse.ac.uk/103070/1/Kriticos_enhancing_property_tax_compliance_ Mandalay_published.pdf

Blöchliger, H. 2015. Reforming the tax on immovable property: Taking care of the unloved. OECD Economics Department Working Papers. 1205. 1815-1973.

Bolc, L. and Cytowski, J. 1992. Search Methods for Artificial Intelligence. Academic Press. Borst. R. 2020. "An AutoML Experiment." internal report, Tyler Technologies Inc.

Bozinovski, S. 2020. "Reminder of the First Paper on Transfer Learning in Neural Networks, 1976."

Informatica. 44(3): 291-302.

Breuker, J. and Van de Velde, W. 1994. CommonKADS Library for Expertise Modelling: Reusable Problem Solving Components. IOS Press.

Crommelinck, S. et al. 2019. “Application of Deep Learning for Delineation of Visible Cadastral Boundaries from Remote Sensing Imagery.” Remote Sensing. 11(21): 2505. Found at: https://www.mdpi. com/journal/remotesensing/special_issues/Land_Administration.

Dasso, J. 1973. Computerized Assessment Administration. IAAO.

Elsinghorst, S. 2018. “Machine Learning Basics Gradient Boosting & XGBoost.” Found at: https:// www.shirin-glander.de/2018/11/ml_basics_gbm.

Fountaine, T. et al. July 2019. “Building the AI-Powered Organization” Harvard Business Review. 62-73. Found at: https://hbr.org/2019/07/building-the-ai-powered-organization.

Friedman, J. 1999. “Greedy Function Approximation: A Gradient Boosting Machine.” Found at: https://statweb.stanford.edu/~jhf/ftp/trebst.pdf.

Goodfellow I.; Bengio, Y. and Courville, A. 2016. Deep Learning. The MIT Press.

Google. 2020. “k-Means Advantages and Disadvantages | Clustering in Machine Learning.” Last modified February 2, 2020: https://developers.google.com/machine-learning/clustering/algorithm/ advantages-disadvantages.

Gorunescu, F. 2011. Data Mining: Concepts, Models and Techniques. Intelligent Systems Reference Library, vol 12. Springer. Found at: https://link.springer.com/book/10.1007/978-3-642-19721-5.

Graves, A.; Mohamed, A. and Hinton, G. 2013. “Speech Recognition with Deep Recurrent Neural Networks.” 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vancouver, British Columbia, Canada. 6645-,649. Found at: https://ieeexplore.ieee.org/ document/6638947.

Han, J.; Kamber, M. and Pei, J. 2012. Data Mining: Concepts and Techniques. 3rd ed. Waltham, Mass.: Morgan Kaufmann Publishers. Found at: http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/ The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-PeiData-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf.

Hastie, T.; Tibshirani, R. and Friedman, J. 2017. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer. Found at: https://web.stanford.edu/~hastie/ElemStatLearn/ printings/ESLII_print12.pdf.

Hochreiter, S. and Schmidhuber, J. 1997. “Long Short-Term Memory.” Neural Computation. 9(8): 1735-1780.

Holland, J. 1975. Adaptation in Natural and Artificial Systems. University of Michigan. Holland, J. 1992. Adaptation in Natural and Artificial Systems. 2nd ed. The MIT Press.

International Property Tax Institute. 2021. IPTI White Paper: “The Potential of Artificial Intelligence in Property Assessment.” Found at: http://ipti.org/wp-content/uploads/2021/08/Potential-of-Artificial-Intelligence-in-Property-Assessment-IPTI-White-Paper-Final-26-July-2021.pdf.

James, G. et al. 2013. An Introduction to Statistical Learning: With Applications in R. Springer. Found at: https://link.springer.com/content/pdf/10.1007%2F978-1-4614-7138-7.pdf.

Jeffrey, R. 1981. Formal Logic: Its Scope and Limits. McGraw-Hill.

Jensen, D.L. 2013. “Tricks of the CAMA Masters.” Journal of Property Tax Assessment & Administration. 10(4): 39-67.

Jensen, D.L. 2015. “Tricks of the CAMA Masters II.” Journal of Property Tax Assessment & Administration. 12(2): 95-124.

Krizhevsky, A.; Sutskever, I. and Hinton, G. 2012. “ImageNet Classification with Deep Convolutional Neural Networks.” In Advances in Neural Information Processing Systems, eds.

F. Pereira et al. Curran Associates. Found at: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Kuhn, M. and Johnson, K. 2013. Applied Predictive Modeling. Springer.

LeCun, Y. 1989. “Generalization and Network Design Strategies.” Technical Report CRG-TR-89-4, University of Toronto.

LeCun, Y. et al. 1998. “Gradient-Based Learning Applied to Document Recognition.” Proceedings of the IEEE. 86(11): 2278-2324.

Lowrance, R.E. 2015. “Predicting the Market Value of Single-Family Residential Real Estate,” PhD thesis. New York University.

McCulloch, W.S. and Pitts, W. 1943. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” Bulletin of Mathematical Biophysics. 5(4): p. 115-133.

Mnih, V. et al. 2013. “Playing Atari with Deep Reinforcement Learning.” arXiv. Found at: https://arxiv. org/pdf/1312.5602.pdf.

Muhammad, M.S. and Ishiyaku, B. 2013. “An Assessment of the Prospects of Property Tax Administration in Nigeria: A Case Study of Bauchi State Board of Internal Revenue.” Elixir International Journal of Social Science, 59, 15284-15289.

Neene, V. and Kabemba, M. 2017. “Development of a Mobile GIS Property Mapping Application Using Mobile Cloud Computing.” Journal of Advanced Computer Science and Applications. 8(10): 57-68.

New York City Mayor’s Office of the Chief Technology Officer. 2021. “AI Strategy: the New York City Artificial Intelligence Strategy.” Found at: https://www1.nyc.gov/assets/cto/downloads/ai-strategy/ nyc_ai_strategy.pdf.

Olatunji, O.C. and Ayodele, K.B. 2017. “Impact of Information Technology on Tax Administration in Southwest, Nigeria.” Global Journal of Management and Business Research: D Accounting and Auditing 17(2).

Oluwadare, C .O. and Ojo, O. 2014. “Database Creation for Tenement Rate Collection: The Role of GIS.”

Pacific Journal of Science and Technology. 15(1): 190-199.

Park, S. and Song, A. 2020. “Discrepancy Analysis for Detecting Candidate Parcels Requiring Update of Land Category in Cadastral Map Using Hyperspectral UAV Images: A Case Study in Jeonju, South Korea.” Remote Sensing. 12(3): 354. Found at: https://www.mdpi.com/journal/remotesensing/ special_issues/Land_Administration.

Perron, L. and Furnon, V. 2019. OR-Tools 7.2. Found at: https://developers.google.com/optimization/.

Rosenblatt, F. 1958. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.” Psychological Review. 65(6): 386-408.

Russell, S. and Norvig, P. 2020. Artificial Intelligence: A Modern Approach (4th ed.). Pearson.

Schwefel, H.P. 1965. “Cybernetic Evolution as a Strategy for Experimental Research in Fluid Mechanics.”

Master’s thesis (in German). Hermann-Fottinger Institute for Fluid Mechanics, Technical University of Berlin. Found at: https://ls11-www.cs.tu-dortmund.de/people/rudolph/publications/papers/

HPS-Rudolph.pdf

Schwefel, H.P., 1995. Evolution and Optimum Seeking. Wiley Press. Sejnowski, T. J. 2018. The Deep Learning Revolution. The MIT Press.

Sowa, J.F. 1983 Conceptual Structures: Information Processing in Mind and Machine. Addison-Wesley. Ullman, J. 1988. Principles of Database and Knowledge-Base Systems, Vol. 1. Computer Science Press.

Vaswani, A. et al. 2017. “Attention Is All You Need.” In Advances in Neural Information Processing Systems, eds. I. Guyon et. al. Curran Associates. Found at: https://papers.nips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Wehrli, J. 2019. “Leveraging Optimization and Constraints in Variable Selection.” 85th International IAAO Conference. Found at: https://researchexchange.iaao.org/conference/IAAO2019/schedule/76/.

Wilson, J. and Daugherty, P. July-August 2018. “Collaborative Intelligence: Humans and AI Are Joining Forces.” Harvard Business Review. Found at: https://hbr.org/2018/07/ collaborative-intelligence-humans-and-ai-are-joining-forces.

Zurada, J.; Levitan, A. and Guan, J. 2011. “A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context.” Journal of Real Estate Research 33(3): 349-388.