

Artificial Intelligence and Machine Learning: Current Applications in Real Estate

by

Jennifer Conway
B.A. Architecture, 2010
University of California, Berkeley

Submitted to the Program in Real Estate Development in Conjunction with the Center for Real Estate in
Partial Fulfillment of the Requirements for the Degree of Master of Science in Real Estate Development

at the

Massachusetts Institute of Technology
September 2018

©2018 Jennifer Conway
All rights reserved

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic
copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author:

Center for Real Estate
July 27, 2018

Certified by:

Alex Van De Minne
Researcher, Department of Urban Studies and Planning
Thesis Supervisor

Certified by:

David Geltner
Associate Director of Research, Center for Real Estate,
Professor of Real Estate Finance, Department of Urban Studies and Planning
Thesis Supervisor

Accepted by:

Professor Dennis Frenchman
Class of 1922 Professor of Urban Design and Planning,
Director, Center for Real Estate
School of Architecture and Planning

THIS PAGE INTENTIONALLY LEFT BLANK

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING: CURRENT APPLICATIONS IN REAL ESTATE

by

Jennifer Conway

*Submitted to the Program in Real Estate Development in Conjunction with the
Center for Real Estate in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Real Estate Development*

Abstract

Real estate meets machine learning: real contribution or just hype? Creating and managing the built environment is a complicated task fraught with difficult decisions, challenging relationships, and a multitude of variables. Today's technology experts are building computers and software that can help resolve many of these challenges, some of them using what is broadly called artificial intelligence and machine learning. This thesis will define machine learning and artificial intelligence for the investor and real estate audience, examine the ways in which these new analytic, predictive, and automating technologies are being used in the real estate industry, and postulate potential future applications and associated challenges. Machine learning and artificial intelligence can and will be used to facilitate real estate investment in myriad ways, spanning all aspects of the real estate profession -- from property management, to investment decisions, to development processes -- transforming real estate into a more efficient and data-driven industry.

THESIS SUPERVISORS:

Alex Van De Minne, Researcher

David Geltner, Professor of Real Estate Finance and Associate Director of Research

Acknowledgements

This thesis was a product of many brilliant minds who have influenced my education and research interests and created incredible companies that are quickly changing the world.

It would not have been possible without my thesis advisors Alex Van De Minne and Professor David Geltner. Alex was a grounding force in my research, teasing out an academic approach and thoughtful considerations of the implications of this topic. David was encouraging, thought provoking and, as always, helpful in getting to the bottom of what really matters in real estate investment.

I am grateful to many staff and faculty at the MIT Center for Real Estate including, but not limited to, Tricia Nesti, our steadfast answerer of all questions; Professor Albert Saiz, who inspired us all to write excellent “books” for this project; Jennifer Cooke, my most excellent academic advisor; Dr. Andrea Chegut, real estate scientist extraordinaire; and Steve Weikal, who encouraged me to pursue this topic and made some key industry connections.

My thanks also go to friends, family, and colleagues who helped with this research in a variety of ways, including Ali Faghah, Alex Conway, Michael Beckerman, Sara Shank, Shaw Lupton, Tony Mustoe, and Ra’ad Siraj. I would also like to thank many others who offered encouragement and advice along the way.

The incredible support from everyone at MITIMCo has been especially important to my success over the past 4 years while I studied at MIT. I cannot imagine pulling this off if my coworkers and the management team had not been behind me all the way. I would like to particularly thank Michael Owu, my mentor and supervisor since arriving in Boston in 2012, who always pushes and challenges me to attain bigger and better goals.

My love and thanks goes to my friends and family, who have been instrumental to my education and career pursuits for many years. My mother, Sarah Mustoe, is my editor, teacher, companion, supporter, and cherished friend. My father, John Conway, inspires in me a love for science, computers, and learning. My brother, Alex Conway, fuels my creative and academic interests with humor and intelligence. My partner, Mr. Alejandro Elias Viriato, gives me the confidence to pursue perfection.

Contents

Executive Summary	7
Introduction	12
Chapter 1 : Methodology.....	14
Chapter 2 : Overview of Artificial Intelligence and Machine Learning.....	17
2.1 Definitions	17
2.1.1 Big Data	17
2.1.2 Artificial Intelligence	18
2.1.3 Machine Learning	19
2.1.4 Deep Learning	24
2.2 Model Requirements	25
2.3 Machine Learning Shortcomings and Pitfalls	26
2.4 Conclusion	28
Chapter 3 : AI and ML in Commercial Real Estate Today.....	29
3.1 Introduction	29
3.2 Real Estate Tech	29
3.3 Real Estate AI and ML	31
3.3.1 Data Gathering and Distribution	35
3.3.2 Analytics	38
3.3.3 Automated Valuation Models	39
3.3.4 Risk Assessment	40
3.3.5 Business Processes	41
3.3.6 Natural Language Processing/Natural Language Generation	43
3.3.7 Computer Vision/Image Processing	44

3.3.8 3D Augmentation and Space Planning	46
3.3.9 Geospatial Analytics	47
3.3.10 Internet of Things	49
3.4 Conclusion	50
Chapter 4 : Opportunities for Future Applications	53
4.1 Introduction	53
4.2 Case Study: Data and Analytics, Computer Vision, 3D Augmentation, and Geospatial Analytics for Urban Planning	54
4.3 Case Study: Business Processes, Data and Analytics, and 3D Augmentation and Space Planning Enhancements for Construction Cost Estimating	56
Chapter 5 : Concluding Remarks	62
Appendix A - Algorithms	64
Appendix B - <i>Real Estate Technology Company Case Studies</i>.....	77
Glossary	112
Works Cited	113

Executive Summary

Real estate meets machine learning: real contribution or just hype? This is the question many real estate executives and technology firms are trying to answer in today's quickly evolving world of technological advancement. Creating and managing the built environment is complicated and fraught with difficult decisions, challenging relationships, and a multitude of variables. Technology experts are building computers and software that can help resolve many of these challenges, some of them using what is broadly called artificial intelligence (AI) and machine learning (ML).

Artificial intelligence (AI) is a general term for machines performing tasks that typically require human intelligence. A wide variety of applications could fall under this umbrella, but as technology advances our expectations for what computers should be able to automate and augment keeps growing. Machine learning (ML) allows us to learn from the past to predict the future by leveraging "big data" through data aggregation, tracking, analytics, and more. The burgeoning applications of machine learning are bringing the world closer to true AI, and real estate is no exception.

In this thesis I define machine learning and artificial intelligence for the investor and real estate audience, examine the ways in which these new analytic, predictive, and automating technologies are being used in the real estate industry, and postulate potential future applications and associated challenges.

Research Methodology

This thesis begins with a high-level overview of machine learning and artificial intelligence informed by a review of textbooks and online resources. With this high-level understanding of artificial intelligence, machine learning, and some of the techniques available to data scientists, I was able to conduct interviews and online research to identify real estate technology companies that use machine learning or artificial intelligence and seek answers to two main questions:

- A. How is AI/ML being applied in real estate technology today?
- B. What are the actionable opportunities for applications of AI/ML in real estate?

With the answers to these questions and the specifics on applications, commercial real estate investors can be better equipped to assess technology solutions and to capture the value of their own data. Real estate technology companies and entrepreneurs can leverage an understanding of the tools that exist as well as some potential areas to target for future innovations.

In order to answer these questions I conducted in person and phone interviews with over 30 industry professionals and researchers including technology company founders, data scientists, professionals working on technology for large real estate firms, and other industry professionals. I also gathered information on real estate technology firms and trends from company websites and other publications such as trade journals, articles, and reference websites. In addition, I reviewed lists from several sources with thousands of real estate technology firms to identify those that use machine learning or AI in some way. In order to understand how these tools are being applied to the industry I also categorized companies by real estate functional areas described in Table 1.

TABLE 1

REAL ESTATE FUNCTION	FUNDING	NUMBER OF FIRMS	DESCRIPTION
BROKERAGE AND SALES	63%	16	Platforms and services for the real estate sales and leasing process
BUILDING OPERATIONS & PROPERTY MANAGEMENT	12%	21	A range of services and software for managing and operating buildings or tenant spaces
CONSTRUCTION/AE	0%	4	Construction, design, and engineering industry tools
CROWDFUNDING	0%	0	Project funding through large groups of people online
DATA & ANALYTICS	7%	12	General data and analytics providers for real estate
DEVELOPMENT	3%	5	Platforms or services specific to development investors and underwriters
INCUBATOR	0%	0	Collaborative programs to help startups succeed
LEGAL/CONTRACTS	1%	6	Real estate law and contract services
LENDING	1%	3	Services and tools for the lending/mortgage industry
LOCATION/GEOSPATIAL	2%	4	Provide services or products related to geography or mapping
SPACE-AS-SERVICE	0%	0	Non-traditional space leasing, typically with short durations and services from the landlord
VALUATION	0%	3	Tools, including AVMs, to evaluate current values and project future values
OTHER	10%	6	Other tools that may serve broader industries, such as cyber security, urban farming, insurance, and parking
TOTAL	100%	80	

Findings

Based on the interviews and my research it is apparent that machine learning is being applied in myriad ways to various parts of the real estate industry, including property management, valuations, leasing, and construction. 80 companies were identified as real estate technology firms that use machine learning or artificial intelligence techniques for their applications. Table 1 summarizes the amount of funding and the number of firms within each functional area that use machine learning and artificial intelligence.

The types of machine learning and artificial intelligence used in real estate vary greatly. Many companies use multiple forms of machine learning or use it for a variety of purposes. The area with the most focus is data, referring to companies that use machine learning in some way to gather, merge, simplify, or generate data. Many of these firms are not yet at the point of using machine learning analytics to apply data to problems or to find actionable insights, but many firms have ideas for how that can be done in the future. One way will be through advanced analytics using machine learning techniques, which is

already a major component of the real estate technology landscape, but has lots of room to grow. Other forms of machine learning and artificial intelligence that are particularly relevant to real estate include automated valuation models, risk assessment for lending and insurance, business process enhancement through AI tools, natural language processing to understand our data and chat, computer vision to assess imagery and video, 3D design and space planning tools, geospatial analytic tools for understanding location impacts, and the internet of things (IOT) that can track our movements and adjust the thermostat. Table 2 outlines the distributions of the types of machine learning used by the 80 companies identified in this research.

TABLE 2

MACHINE LEARNING TECHNIQUES	COMPANIES	PERCENTAGE
DATA GATHERING/DISTRIBUTION	58	73%
ANALYTICS	53	66%
VALUATION	17	21%
RISK ASSESSMENT	7	9%
BUSINESS PROCESS	4	5%
NATURAL LANGUAGE PROCESSING/GENERATION	16	20%
COMPUTER VISION	14	18%
3D/SPACE PLANNING	13	16%
GEOSPATIAL ANALYTICS	28	35%
INTERNET OF THINGS (IOT)	14	18%

These techniques are being used in interesting and unique ways by each firm. Some of the ways in which these tools add value for real estate are described below.

Data gathering and distribution can be accomplished with new analytics tools, which have the power to capture every interaction with a new software, every sensor output, and every message sent, making these and other day-to-day elements of the real estate business into rich sources of data. The hope is that in the right hands, ML and AI will open the door to documenting more data points, and leveraging existing data for actionable insights.

Analytics using machine learning has great potential to enhance our understanding of the built world and investment. It seems that many real estate technology firms are still in the data gathering phase of their growth. However, as more users join their platforms and these datasets grow, it will be important to track which companies begin to develop truly sophisticated data analysis methods to leverage that data.

Real estate **valuation** is key to the real estate business in areas such as sales, portfolio management, REIT valuations, tax assessment, and lending. Automated valuation models are gaining in popularity and machine learning tools used for this purpose appear to improve predictive abilities, and also streamline the appraisal and assessment process.

Machine learning is a useful tool for **risk assessment** as it can be used to process large data sources and get a better understanding of thousands of potential risk factors, including those that may not have been considered in the past due to limits on data analysis capabilities. With machine learning tools lenders can optimize borrowing levels and rates and insurance companies can balance customer demand and potential costs.

Business processes can be enhanced through new software offerings that by create structured environments in which to conduct business practices. These environments can become data sources themselves, recording and learning from every customer interaction.

Natural language processing is a powerful tool that allows computers to “read” text and extract pertinent data. **Natural language generation** give computers to ability to communicate back to us in “human” terms. These tools are used in real estate related chat bots, contract review and data extraction, data gathering and processing from text based sources, and document writing.

With the powerful tools using **computer vision** technology, which allows computers to interpret images and video, along with the processing power of modern computers, there are many possibilities for how applications in real estate. These include value assessments and quality markers from imagery of buildings, people and vehicle tracking, and reading documents using optical character recognition. The maxim that “a picture is worth a thousand words” is becoming truer from the perspective of real estate investors.

AI and ML technologies are enabling a wide range of new approaches to mapping, designing, and constructing the built world. With machine learning technology, **3D augmentation and space planning** can be performed quickly and efficiently.

In order to analyze real estate within its physical context it is useful to employ **geospatial analysis**, “the gathering, display, and manipulation of imagery, GPS, satellite photography and historical data, described explicitly in terms of geographic coordinates or implicitly, in terms of a street address, postal code, or forest stand identifier as they are applied to geographic models” (“Definition - What is geospatial analysis?,” n.d.). The importance of location in real estate is not lost on the machine learning scientist who have develop many tools that integrate with geospatial information to help us better understand locational relationships and predict outcomes with this important layer of information in the mix.

With the growth of massive datasets coming from **Internet of things (IoT)** technology, where physical objects in our environment are internet connected, we will be able to see patterns in human behavior and interactions with the physical world that we have never seen before. This opens up possibilities for solving all kinds of problems, from the mundane need to ensure comfortable temperatures in a room, to global issues such as health impacts of air quality, or the spread of disease.

The implications of these various techniques and applications are far reaching in the real estate world, but we are still in the early days of developing and perfecting their use. The data generated by technology firms of all kinds could be better leveraged through machine learning to provide even more tools and insights, and eventually evolve into a more general AI for real estate.

Observations

There are many hurdles to overcome in applying machine learning to real estate, including data quality and inherent challenges associated with real estate data. For example, many machine learning techniques are not great for analyzing time series data, which is very important for real estate investment trends. However, the biggest challenge may be persuading the professionals who will make decisions about software and data management to get on board. Real estate professionals have long been slow to adopt new technologies. Professor David Geltner observed that as in past technology

trends, firms seem reluctant to adopt AI and ML due to inertia, cultural frictions, a desire to not share data, avoidance of accountability, and vested interests in the way things are done today (Geltner, 2018).

Getting high-quality data is the first step to applying machine learning, but it can be difficult to find in an industry that has long suffered from siloed processes, considers most data proprietary, and is plagued by numerous errors and inconsistencies in the collection and recording of data. Over 70% of the companies entering the machine learning space for real estate are in some way focused on gathering, cleaning, merging, or creating data. Once more data has been generated and aggregated more doors will be open for advanced analytics and creative uses of the data itself. Companies that are actively using machine learning in areas unrelated to data gathering and aggregation are still in very early stages. Often, a human verification stage is still required for many processes or analyses, or in the case of automated valuation models, the results may be checked against those of other models for accuracy and reliability.

Real estate professionals may be reluctant to adopt or trust technology, and it will take true results with evidence of ROI to gain market share. In spite of this hurdle, there are many companies that are thinking about ways to leverage machine learning now that the data is coming in, and small but important insights about the real estate business are beginning to emerge. Real estate experts are taking note and beginning to invest in understanding, and even promoting these moves. As real estate professionals incrementally start to adopt machine learning in small areas of the business it can steadily grow to become an important part of the industry.

An additional observation was that employees or founders with a strong background in machine learning are key to making ML and AI primary and useful components of the various applications. Machine learning and artificial intelligence are complex areas requiring careful consideration of many variables. It will be important for real estate technology firms to attract the best talent in machine learning and artificial intelligence in order to progress in these areas, but there is a lot of competition in this area and scarcity of programmers and data scientists.

There is a lot of hype about the promise of AI and analytics powered by machine learning, but there are also some truly remarkable efforts going on that will eventually change the ways in which we do business. There are also areas where we can think bigger in the longer term about how to capture the most value from these tools. For example, the field of urban planning could be transformed by data gathered using new sensor technologies with a focus on human interactions with the spaces they inhabit. There could also be transformations to the way we build our environment; what is today an inefficient and costly endeavor could become automated and streamlined to produce better, faster, safer results from construction projects. All of this will take time but some form of this evolution is inevitable.

Technology moves quickly and it may seem difficult to keep up, but, if we can see past the excitement of media attention, there are plenty of examples of ground-breaking research that is truly opening up insights that never would have been possible before the advent of AI and ML. I expect that the real estate technology landscape will evolve swiftly and it will become ever more challenging to track the companies and research that is changing the way we do business.

Introduction

Artificial intelligence (AI) and machine learning (ML) technologies have the potential to transform the way our world works. This is a commonly heard aphorism in today's quickly evolving technology world. It seems likely that over time these tools and other related technologies will indeed be transformative, but we have a long way to go and many lessons to learn along the way. The real estate investment world, and particularly commercial real estate, has been slow to adopt new technologies, including those that utilize machine learning, and we have yet to see true implementation of artificial intelligence. However, some emerging technology and data firms harness the value of these technologies for a wide array of uses in the real estate industry. This thesis explores two questions through the lens of a commercial real estate investor and technology enthusiast:

- C. How is AI/ML being applied in real estate technology today?
- D. What are the actionable opportunities for applications of AI/ML in real estate?

Interviews conducted with more than 30 real estate and technology professionals and researchers provided insight into these questions. Combining with information gathered from real estate technology company websites and other publications, I attempt in this paper to present a holistic picture of machine learning and artificial intelligence in real estate as it is being used today and plans for use in the near future. The overall finding is that machine learning is being applied in myriad ways to various parts of the real estate industry, including property management, valuations, leasing, and construction. Firms with a strong interest in machine learning and analytics have benefitted from founders and early employees with a strong background in data science and machine learning, which are complex areas requiring careful consideration of many variables involved in developing these types of technologies. However, we are still in the early days of developing and perfecting these technologies. The data generated by technology firms of all kinds could be better leveraged through machine learning to provide even more tools and insights, and eventually evolve into a more general AI for real estate.

With the answers to these questions and the specifics on applications, commercial real estate investors can be better equipped to assess technology solutions and to capture the value of their own data. Real estate technology companies and entrepreneurs can leverage an understanding of the tools that exist as well as some potential areas to target for future innovations.

Context

We are in a new world of data – big data, wide data¹, data visualization, data capture, data analysis techniques, and more. Computers are being taught to take in data points and churn out insights to be used in industries of all kinds. If one looks at any tech publications or attends a tech conference, one encounters statement such as Gartner's claim that "during the next few years, virtually every app, application and service will incorporate some level of AI" (Cearley, Burke, Searle, & Walker, 2017). News reports highlight statistics such as, "machine learning patents grew at a 34% Compound Annual Growth Rate (CAGR) between 2013 and 2017, the third-fastest growing category of all patents granted" (*8 Fast Growing Technologies*, 2018), or "Forrester predicts the Predictive Analytics & Machine Learning (PAML) market will grow at a 21% CAGR" (Columbus, 2018). According to real estate tech company Truss Co-

¹ "wide data" is a phrase coined by Dr. Andrea Chegut to describe integration of many types of data and platforms, such as her lab, the MIT Real Estate Innovation Lab, is doing for NYC data.

Founder Tom Smith, “the types of modeling we can do now are crazy compared to what we could do a few years ago” (2018).

Real estate technology is seeing a boom in interest from investors and real estate professionals. According to Forbes, in 2017, venture investors deployed over \$5 billion in real estate technology, more than 150 times the \$33 million invested in 2010 (Snider & Harris, 2018). Investors like Fifth Wall, a real estate focused VC fund, are taking this huge growth in one of the world’s biggest, but “under-teched” industries as an opportunity to implement and distribute these new technologies (Wallace & Bruss, 2018). We are even starting to see equity investments in “prop tech” coming directly from the real estate industry; property owners are interested in making transactions easier, cheaper, and with less friction, so they are keeping an eye on and supporting technology that can help this along (Smith, 2018).

As the real estate industry works to catch up in the realms of technology and data collection, uses of new machine learning analytics technologies could quickly advance. Huge value can be captured from the reams of data sitting in databases everywhere and being aggregated in new tech platforms, benefitting real estate and technology investors alike. Technology VC’s are taking note. In a press release for a recent funding commitment, Sequoia Capital partner Haim Sadger said “The promise of AI to transform commercial real estate investments cannot be [over]stated. Over the last few years, we’ve seen AI disrupt a number of traditional industries and the real estate market should be no different” (Shu, 2018). According to CRE Tech founder Michael Beckerman, AI is new to the real estate industry, but it is coming on fast” (Beckerman, 2018). Real estate is one of the fields where big data, machine learning, and artificial intelligence are getting a lot of attention, but in spite of the hype the applications to date are limited.

While these trends and statements indicate excitement and vast potential for AI, there are many reasons to question the speed at which change will occur and the extent of those changes. As of July 2017, Gartner placed cognitive computing, machine learning, and deep learning at “The Peak of Inflated Expectations,” heading quickly towards the “Trough of Disillusionment” (Panetta, 2017). There are many roadblocks and pitfalls that will impact the effectiveness and outcomes of these new applications (Prado, 2018). There are also examples of attempting to use machine learning to fix an inherently broken system that would better be addressed through smarter processes (Spitzen, 2018). Furthermore, it is a common truism that lots of people are doing machine learning, but they are doing it poorly.

Machine learning and artificial intelligence is just one piece of the equation for bringing the industry up to speed, but it can have wide-ranging applications in the quickly evolving real estate world.

Outline

In Chapter 1 I provide an introduction to the research methodology for this paper. Next, in order to get everyone on the same page, Chapter 2 will provide an overview of machine learning and artificial intelligence as they will be defined in this paper. Chapter 3 introduces the real estate technology companies that are currently using machine learning or artificial intelligence methodologies in some capacity and present use cases in the real estate industry using these specific examples. Chapter 4 presents ideas on how machine learning and AI technology could be employed in the future to enhance and evolve the real estate industry. Additional material on specific machine learning algorithms is provided in Appendix A, with brief definitions and diagrams. Appendix B includes details on many of the companies interviewed or researched for this paper.

Chapter 1 : Methodology

The research for this thesis was conducted through online research, academic literature review, and one-on-one interviews with real estate technology entrepreneurs, researchers, and investors.

There are various sources of information on technology firms such as AngelList, CB Insights, and Crunch Base. These track many attributes, including location, funding level, founding date, contact information, websites, and more. There are also several sources that are specific to real estate technology such as CRE Tech (cretech.com). I compiled a list of over 2,000 real estate technology firms using resources from realestatetech.co, CB Insights, CRE Tech, MIT Real Estate Innovation Lab, personal interviews, and various media sources. Based on a review of descriptions, press, and websites, 80 companies were identified as using some form of artificial intelligence or machine learning in their platforms. Due to the large number of technology firms focused on real estate and the difficulty in identifying where ML and AI are used, this list may exclude some firms. It is not intended to be an exhaustive list as new firms are joining the ranks of prop tech all the time, but these examples are indicative of market trends and capture many of the most prominent and promising players.

In order to get a better understanding of how real estate technology firms employ machine learning and AI ideas and techniques I conducted interviews with over 30 people. Twenty one of these interviews were with founders, CEOs, CTOs, data scientists, engineers, sales team members, and other employees of technology firms of various types and sizes. For example, I spoke with established data and analytics providers RCA, based in New York, and Ortec Finance, based in the

Netherlands. I also connected with data startups like Foxy AI, Locate AI, and Enertiv, and with young analytics and valuation firms including Enodo, REview, and Navigator CRE. Other startups were more focused on providing a platform or service, such as Leverton's contract abstracting, Mynd's property management services and Truss's leasing platform. For each interview I asked the following questions in some form:

- What area of technology/real estate are you/your company focused on?
- I am interested in understanding how machine learning is being used in real estate today. Is this something that you/your company are using? If so, please describe how the techniques are being applied.
- Are you aware of any other ways that machine learning is being applied in real estate or related fields today?
- Are there any areas of the real estate business that you think would be good opportunities for applying machine learning or artificial intelligence techniques in the future?
- Many articles and media outlets are talking about machine learning and AI, but there may be inconsistencies in what is considered machine learning or AI. How would you define machine learning and artificial intelligence?
- Data is important to machine learning and AI modeling and tools. What are the key data sources in your business? What types of data are you using/collecting to be used in your product or service? What types of data might be missing or useful to enhance analysis or machine learning and AI applications in real estate?

Each firm had varying degrees of concern over privacy regarding their technology and business practices, so the material presented here

mostly reflects general principles of how to use machine learning and the types of problems being solved rather than detailed information on algorithms used.

I also spoke to industry experts Yishai Lerner, a Co-CEO at JLL Spark, Jones Lang Lasalle's new technology investment fund; Natalie Bruss and Brendan Wallace of Fifth Wall Ventures; Elie Finegold, real estate technology entrepreneur; and Chandra Dhandapani, Chief Digital & Technology Officer at CBRE. I discussed with them where they are investing their resources, the technologies they expect to be most promising, and how machine learning and artificial intelligence could be applied in real estate. They were able to provide valuable insight into the types of technologies and companies they are tracking and how they approach real estate technology investment and implementation.

In addition, academic papers and textbooks were used as references in order to understand and define machine learning and artificial intelligence. While the mathematical and more technical language has not been used in this paper, the foundational principles on which statistics and machine learning rely are important to understand in relation to their applications in real estate.

The diagrams and tables presented throughout the paper were created for the sole purpose of this thesis, in some cases using the research papers and textbooks referenced as inspiration.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2 : Overview of Artificial Intelligence and Machine Learning

"Those who rule data will rule the entire world."

— Masayoshi Son, Founder/CEO, SoftBank

In this chapter I will define big data, artificial intelligence, machine learning, and associated terms as I intend to use them in this paper. Specifics on selected machine learning algorithms can be found in Appendix A. A glossary is provided at the end of the paper to define terms that appear in italics throughout the paper.

2.1 Definitions

Artificial intelligence is a term heard and seen more frequently in various forms of media all the time. Artificial intelligence captures the imagination with images of robots that act and talk like humans as seen in science fiction over many decades. Machine learning sounds like a complex component of unintelligible computer jargon. So it is no wonder that there is a lot of grey area when companies claim to be using machine learning and artificial intelligence in their businesses; these terms are misused and misrepresented. Therefore, our first step is to define them for the purposes of this research and introduce more precise terminology to describe the technologies.

2.1.1 Big Data

The term *big data* has been thrown around a great deal with little understanding of the meaning, but it is a driving factor that in many cases makes machine learning a powerful tool. Big data here will refer to large data sets, meaning many observations, with many attributes or variables. Big data is often described in terms of the challenges associated with the “3 Vs” -- volume, velocity, and variety -- of data being produced with new technologies, an idea commonly attributed to Gartner analyst Doug Laney. In recent years, discussions have also included veracity and value as additional challenges of dealing with data. The focus of this paper will be on how ML technologies can help gather, manage, clean, and analyze large datasets.

There are many examples of datasets in real estate that could qualify as big data and be candidates for advanced analytics. One example might be a dataset of asset-level information for every property in a city including price, size, and location, but also geospatial metrics, demographics, flood plains, retail sales volumes, rents, hotel rates, and more. This could add up to millions of data points for a moderately sized city.

The number of attributes tracked and the large volume of assets makes this dataset “big.” Another example would be the data produced when using sensors, cameras, and mobile devices to track movements and habits of people throughout a shopping center. This dataset could exist thanks to new sensor, image processing, and mobile tracking technologies.

A common issue raised when discussing real estate analytics is that much of the data is “unstructured,” meaning there is a lack of systematic and consistent ways to track data and the text-heavy nature of the data, whether in the form of property descriptors or deal terms, such as lease terms or construction contracts. Figuring out how to organize unstructured data and use it to make meaningful insights is a major challenge faced in applying machine learning to real estate, but most companies don’t have the capabilities to do it (Hong, 2018). However, with improved analytics capabilities the structure of the data becomes less important because it can be modeled (Grimes, 2005). Machine learning can be an important tool in processing unstructured data to find patterns and create consistency, though this is still a major challenge for data scientists.

As Bishop notes in “Pattern Recognition and Machine Learning,” searching for patterns in datasets is nothing new (2006, p. 1). But with advanced machine learning algorithms and increased computing power, we are now able to analyze more data more quickly. Companies like Google and Facebook have already shown the value of big data. As this trend continues, data will become an intangible asset on company datasheets, especially if it can be structured (Somani, 2018). However, the power of “big data” is not in the data itself, but in what you do with it.

2.1.2 Artificial Intelligence

The term artificial intelligence has been used for decades and has carried different meanings and connotations. A simplistic definition is: systems

that perform tasks that normally require human intelligence. What this means and where we are in terms of technology to achieve this is a moving yardstick; the industry makes progress, but the measure for success continues to move forward (Chegut, 2018). One view is that “the primary focus of AI is on *acting* rather than *thinking*, and on *doing the right thing* rather than *emulating humans*” (Everitt & Hutter, 2018). Computers or machines don’t have to function in a human way with human behaviors to achieve the desired result.

Examples of emerging real estate technologies that could be defined as artificial intelligence include “smart” building operations tools, powered in part by the internet of things (IoT) and that learn resident or tenant preferences for light, temperature, and other space qualities. Another category is document “reading” through natural language processing, including automated lease abstracting. Computer vision and image process is also an area where computers are catching up to humans in the ability to identify and classify objects or scenes.

Gartner, a leading technology research firm, has identified *augmented analytics* as an important form of artificial intelligence for business. This refers to using “machine learning to automate data preparation, insight discovery and insight sharing for a broad range of business users, operational workers and citizen data scientists” (Cearley et al., 2017). With software that can perform complex analytical tasks to achieve actionable insights about data, with little to no human supervision, the analytics capabilities of companies could quickly advance.

However, we are still very far away from contextual, general purpose artificial intelligence in real estate (Chegut, 2018). Computers are not yet able to predict values without hands-on machine learning methods, or call the building operator when a system is down without careful

planning and code writing to enable this to happen. Yet, as demonstrated by a few of the companies examined in this paper, machine learning is opening the door to more automated processes and incremental improvements are coming quickly.

2.1.3 Machine Learning

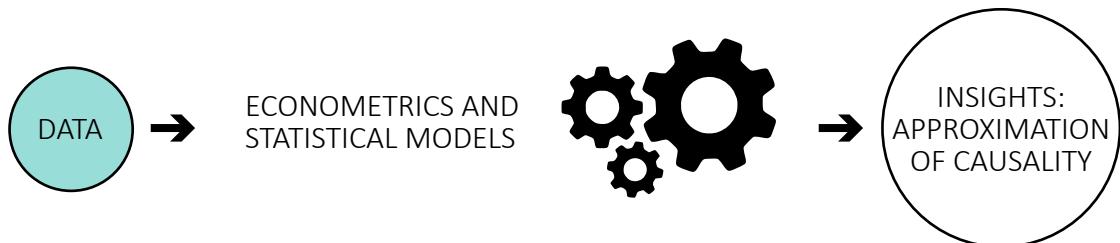
Machine learning (ML) is a type of artificial intelligence and can simplistically be defined as: systems learning from the past to predict the future (“Data Robot AI Experience,” 2018). Algorithms are used to “learn” the relationship between data variables. Another definition states, “an ML algorithm learns patterns in a high-dimensional space without being specifically directed” (Prado, 2018, p. 3). In predictive analytics or predictive modeling those relationships can then be applied to a new data set to predict outcomes. It differs from traditional statistical modeling in that the “rules” are not known beforehand. According to DataRobot, an automated machine learning enterprise provider: “While most statistical analysis relies on rule-based decision-making, machine learning excels at tasks that are hard to define with exact step-by-step rules;” and “machine learning can be applied to numerous business scenarios in which an outcome depends on hundreds of factors that are difficult or impossible for a human to keep track of” (“Machine Learning | DataRobot Artificial Intelligence Wiki,” n.d.). Every statistical model and machine algorithm has its limitations, but there are many use cases where both are used to achieve higher accuracy and more clarity as to the causation of outcomes (Sadler, 2018).

Exhibit 1 is an illustration of the differences in the methodologies. As an example, a traditional valuation model would be implemented by applying a methodology that uses an equation to describe the relationship between asset value and one or more variables. By observing these relationships one can gain a general understanding of the impacts of each variable,

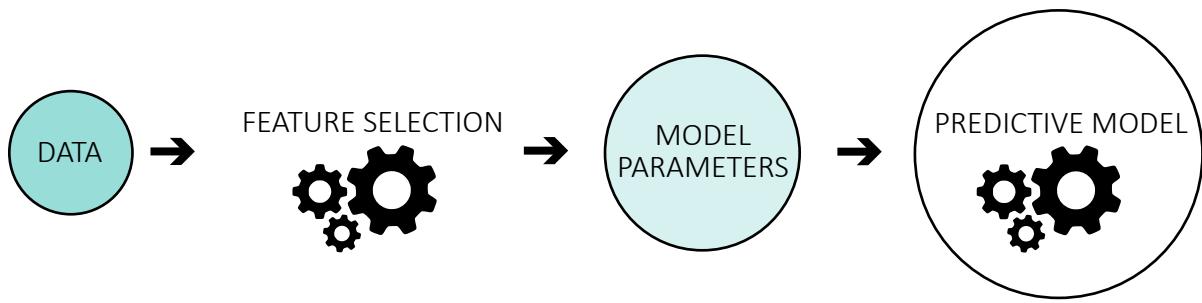
such as location, unit size, and building quality, has on asset value. In a machine learning automated valuation model the key features are automatically identified based on statistical significance in relation to value, and then used to structure a predictive methodology using a progressive set of steps and optimized for precision, complexity, and other factors. Given a set of features of an asset, the model can produce a predicted value, and in some cases an error rate for that prediction.

EXHIBIT 1

Statistical/Econometric Modeling



Machine Learning



In traditional statistical modeling the output is an approximation of causality based on observed relationships in the data.

Machine learning follows a process of selecting the relevant features and takes model parameters that define the trade-offs between precision and stability of the model. Training the model produces a predictive model that can be used to make predictions for unlabeled data.

There are several main categories of machine learning approaches:

Supervised learning: In supervised learning, algorithms describe the relationship between *input variables* or *input vectors* and observed outcomes and applies them to new inputs to predict the outcome. The observed and predicted outcomes are the *target variable* or *target vector*. When the training data includes outcomes, it is sometimes referred to as “labeled data” (Bishop, 2006, p. 2). An example of outcomes could be the interest rate for a mortgage given inputs of property qualities and borrower characteristics.

Supervised problems can further be broken down into classification and regression problems:

- **Classification problems:** This involves breaking data sets into a finite number of buckets or classifications (Bishop, 2006, p. 3). For example, one might split real estate assets into recommendations for the highest and best use types based on known attributes using past examples of HBU studies.
- **Regression problems:** A regression is aimed at giving the value on a continuous spectrum of a target variable based on inputs. An example is predicting real estate asset values given a set of asset qualities and market characteristics.

Unsupervised learning: Unsupervised learning differs from supervised learning in that there are no observed outcomes to start with. Instead the outputs of the model are observations about the data itself, otherwise known as unlabeled data. Unsupervised learning problems include density estimation (determining the concentration of close or similar variables within a dataset), gathering and using internet of things (IoT) data, and data visualization. (Bishop, 2006, p. 3)

Unsupervised learning can be used for both clustering and dimensionality reduction:

- **Clustering:** Used to identify groups of similar data points when there is no defined output

(Bishop, 2006, p. 3). An example in real estate is dividing assets into groups based on physical proximity and building attributes.

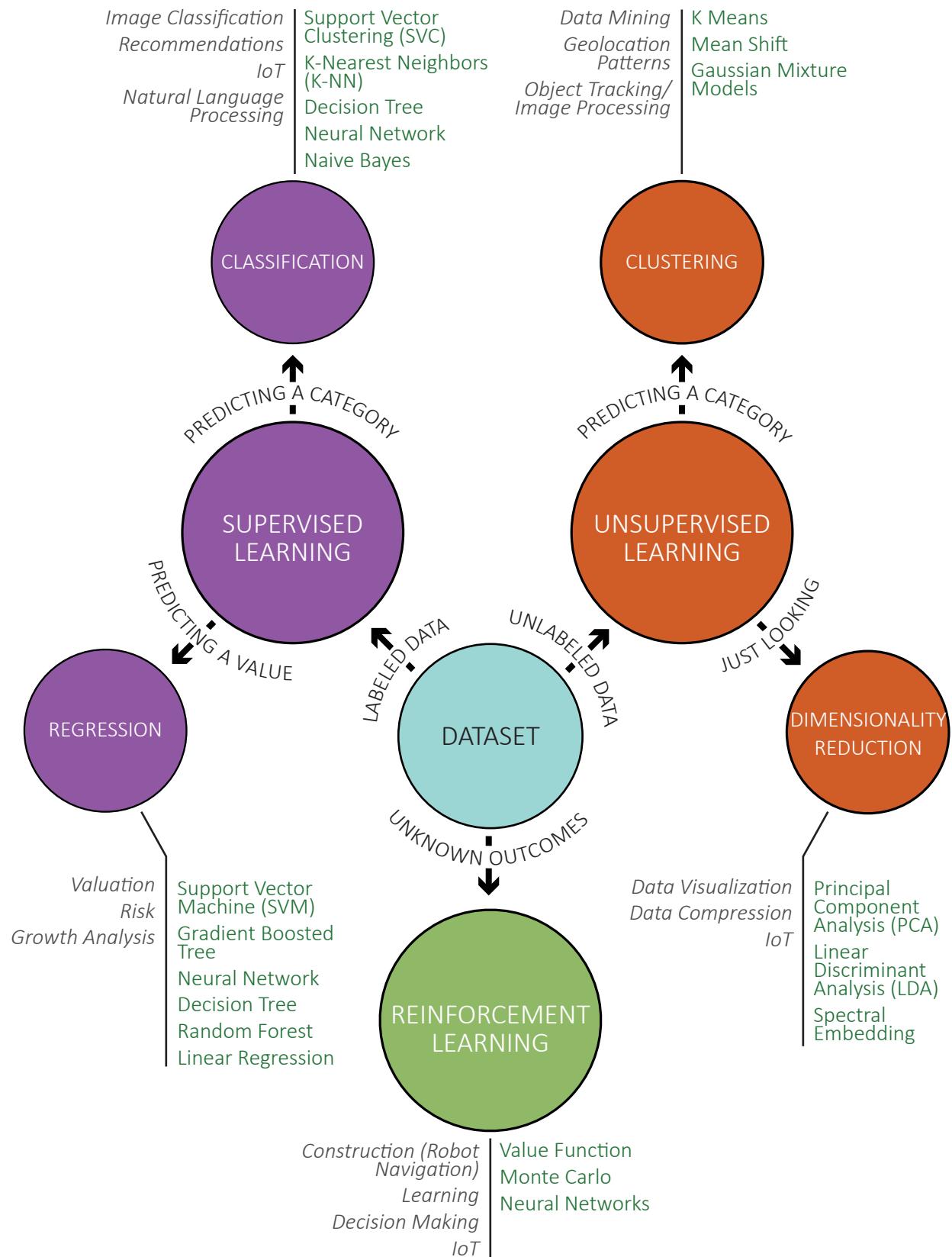
- **Dimensionality Reduction:** This is often a component of data pre-processing, or a component of other machine learning processes to simplify or transform data (Bishop, 2006; “Choosing the right estimator,” n.d.). This can be used in data visualizations such as mapping, and in data compression for large datasets such as those produced by IoT enabled devices.

There are also ways to combine supervised and unsupervised learning, called semi-supervised learning, to achieve better results when only some of the data is labeled data.

Reinforcement learning: In reinforcement learning problems there is no given output but the computer must reach the optimal output by trial and error (Bishop, 2006, p. 3). This type of learning is generally associated with AI and as one example has been used to create systems where computers can play games. In a well-publicized example, in 2016 Google’s AlphaGo beat the best Go player in the world, using its ability to test and “remember” many potential outcomes of millions of different move combinations (Moyer, 2016).

In Exhibit 2, some of the types of machine learning methodologies available are listed in the green text, and the types of applications where they may be used are in italics. This diagram was adapted from Scikit Learn’s algorithm cheat sheet, SAS’s algorithm cheat sheet, and Isazi Consulting’s map of machine learning algorithm types (“Choosing the right estimator,” n.d.; “Isazi Consulting,” n.d.; Li, 2017)

EXHIBIT 2
Machine Learning Types and Applications



As shown in this simplified map of ML, there are many machine learning frameworks out there, described by one tech CEO as a “zoo,” and it can be hard to know which to use, when and where (Sargent, 2018). There are several popular open source machine learning software packages where these frameworks can be found, including TensorFlow, Apache Spark, SciKit-Learn, and Amazon Machine Learning, that can be helpful to data scientists as they attempt to uncover the best methodologies. In addition, new services like DataRobot’s automated machine learning platform can enable testing of tens or hundreds of different algorithms and combinations of algorithms without the tester having data science expertise, otherwise known as the “throwing spaghetti at the wall” method (“Data Robot AI Experience,” 2018).

Typically, in order for a ML model to “learn” relationships within data or predict outcomes it is given a *training set* of data, which is a subset of the overall data available. The remaining data is the *test set* or *holdout set*. Once the model has been trained using the *training set* the results are applied to the holdout set and evaluated for accuracy (Bishop, 2006, p. 2).

In cases where data sets are small and more data is needed, cross-validation methods can be used. The data is broken into sets and the model is run multiple times with a different set being held out each time. The results of these runs are then combined into an average that best describes the entire set. This can be problematic, however, in that performing multiple training runs requires more computing power and the process can result in greater model complexity (Bishop, 2006, p. 33).

Predictive models deal with uncertainty and incorporate probability theory to make predictions, which can be accompanied by the likelihood of accuracy in certain applications. These results then need to be examined through the lens of decision theory in order to translate

them into optimal decision making for the given problem (Bishop, 2006).

Basic descriptions and diagrams of some of the models used in real estate applications are included in Appendix A.

2.1.4 Deep Learning

Deep learning refers to certain types of machine learning, but an exact definition is hard to find. The term was coined around 2006, but various forms have been studied since machine learning began as a field of study (Schmidhuber, 2015, p. 96). It is often associated with *neural networks* where the layers of processing steps are more numerous, or deep, rather than shallow (Schmidhuber, 2015, pp. 86–87). Deep learning is different from other types of machine learning; ML is just predictive, while deep learning is predictive and generative; it can structure unstructured data (Vomero, 2018).

In an article for Nature, researchers provided the following description:

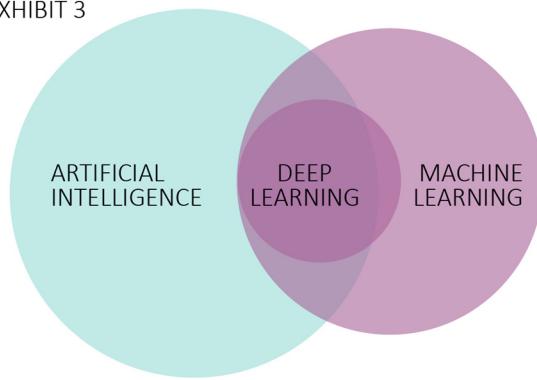
Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer (LeCun, Bengio, & Hinton, 2015).

There are many examples of the applications of deep learning and deep neural networks. One powerful example of a deep neural network that

advances technology to make computers more human is Google Duplex. The platform, announced in May 2018, can be used to have natural sounding phone conversations to complete tasks like making a restaurant recommendation. (Leviathan & Matias, 2018)

Deep learning describes the power of machines to gather and understand data in novel ways, revealing insights never possible before. It is also allowing machine learning to open up possibilities for AI. Exhibit 3 is an illustration of the overlap of all 3 of these concepts.

EXHIBIT 3



This is an imprecise illustration of the overlap of the concepts of AI, ML, and deep learning

2.2 Model Requirements

Those companies that are considering machine learning and artificial intelligence applications for use with big data of various kinds should approach it with their eyes open and goals clearly defined. Jason Mintz of DemystData outlined several factors to consider: the cost of data (time and money), red tape (bureaucracy, politics), and regulatory concerns, and he recommended considering 3 model requirements: usability, predictive power, and explainability ("Data Robot AI Experience," 2018).

While algorithms can be powerful, it is important that the end result be actionable insight into the data (Odegard, 2018). In addition, these insights have to be understood in the context of the problem, asset, or portfolio. There will necessarily be trade-offs between accuracy, stability, simplicity, and interpretation of the model. For example, linear regression has the advantage of producing easily interpretable results, but the accuracy of the predictions are likely to be lower than those achieved using a machine learning technique such as *random forest*. By contrast, while random forest may give very accurate predictions it is challenging to interpret these predictions and understand the reason for which the prediction was made. (Ganes, 2018)

Of course, as with any business decision the ROI should be understood to the extent possible. Adoption will be more likely to happen in an industry that has been slow to advance technologically if real estate professionals see tangible benefits, and look smarter when using it, and it is not necessary that they understand it (Dhandapani, 2018). By starting with targeted use cases, technology firms and real estate professionals will see benefits more quickly than if they try to tackle the big problems with limited hope for early successes (Dhandapani, 2018). Furthermore, if adequate results can be achieved using simple linear regression or other methods,

then implementing machine learning may not pay off. On the other hand, if the data is sufficiently big and the problem complex, then there can be a huge advantage to those who act on the information hidden in their data.

2.3 Machine Learning Shortcomings and Pitfalls

Machine learning and AI are powerful tools, but when used improperly can cause undesirable results. It is important that machine learning be conducted on a basis of sound academic research and real world testing. In every type of algorithm there are various parameters that need to be adjusted to ensure accuracy and validity of outcomes and models with repeatable success.

In a paper titled *10 Reasons Most Machine Learning Funds Fail*, Prado outlines some of the types of machine learning mistakes that can be made in finance, often at great cost. By way of introduction, Prado states, “the rate of failure in quantitative finance is high, and particularly so in financial machine learning. . . . there are ten critical mistakes underlying most of those failures.” He goes on to assert the following:

The flexibility and power of ML techniques have a dark side. When misused, ML algorithms will confuse statistical flukes with patterns. This fact, combined with the low signal-to-noise ratio that characterizes finance, all but ensures that careless users will produce false discoveries at an ever-greater speed. (Prado, 2018)

These are important cautions for data scientists working in real estate to understand.

Nonetheless, ML techniques are powerful and can explain the complexities of modern financial and real estate markets that traditional econometrics cannot: “It is hard to believe that something as complex as 21st-century finance could be grasped

by something as simple as inverting a covariance matrix” (Prado, 2018). Some of the pitfalls and issues with machine learning, specifically in the context of real estate, are discussed below.

Bias-Variance Tradeoff: In any model there needs to be a careful consideration of the *bias-variance trade-off*, where the data scientist has to optimize the level of precision.

In statistical modeling and machine learning over-fitting occurs when the algorithm fits the data with specificity to the given data points, creating an overly complex model that cannot accurately predict results for future observations. For example, if you have ten data points, you can get a perfectly fit model using ten parameters, but then these parameters are basically meaningless, not robust or extensible. This is particularly true when smaller data sets are used; “For a given model complexity, the over-fitting problem becomes less severe as the size of the data set increases” (Bishop, 2006, p. 9). Model complexity can be addressed in multiple ways. For example, some of the parameters can be removed. Another option is to overlay probability techniques and *regularizer* coefficients or functions in some cases, the problem of over-fitting can be addressed (Bishop, 10). A Bayesian model comparison technique moderates the complexity of the model using a function that reduces the influence of each parameter by a ratio determined by the number of parameters in the model (2006, pp. 161–163).

Of course, we do not want to end up with an *under-fit* model; the goal here is to find the optimal balance between the number of parameters and their influence, or, in other words, model complexity.

IID: In many statistical studies an assumption is made about the variables having independent and identical distribution (IID), but can we realistically assume this in real estate? According to Prado, in finance this is typically not the case, for example,

stock prices are going to be highly dependent on the price the day before. It seems reasonable to assume that a similar problem will apply in many real estate problems. Prado offers the *sequential bootstrapping* method as a solution to this issue, at least in finance applications. “We cannot be certain about what observed features caused an effect” (Prado, 2018, p. 11). Since data is not IID, cross-validation techniques do not work when features in both the training and testing set overlap and/or are serially correlated (meaning that the next data point will be close to the one before it because one leads to the next, for example about stock prices over time). To correct for this we can choose testing and training sets that do not overlap (in time) and ensure an appropriately significant gap between them (in time) (Prado, 2018).

Black box: With some machine learning analytics methods the model produced is a “black box” where there is no transparency into how the results are achieved. This has been identified as an issue by businesses, for example, there is a general concept of how a neural network works, but not total insight, which can be a problem, depending on what the customer wants to get out of it (Somani, 2018). In many cases this is not an acceptable condition for business and regulatory reasons. Interpreting deep learning models remains a serious problem and topic of significant ongoing research, as demonstrated by the IBM partnership with MIT to research solutions (Vomero, 2018). According to IBM’s vice president of A.I., Dario Gil, as the lab works to develop new algorithms there is a desire “to move away from the ‘black box’ model, which obfuscates much A.I. research today” (Jones, 2017).

Algorithm purpose: When considering application in real estate it is also important to note that existing algorithms tend to be very general and are not built specifically for real estate, largely due to the sources of those algorithms (Vlaming,

2018). Work and research into the best machine learning algorithms is relatively advanced in some areas of real estate, such as natural language processing for contract abstraction, while others are mostly unexplored or limited to few players, such as the impact of weather and other factors on construction costs. Real estate can borrow from other technologies to inform these new areas of research and practice; however we are still in the early years of the development of these technologies and the applications must be relevant and applied thoughtfully.

Time Series: The development of time series analysis techniques will be one key to certain type of real estate applications, such as valuations. While analytics firms like DataRobot are developing these tools, a challenge is that real estate transactions tend to happen at odd intervals, making it challenging for standard tools to work effectively (Goldstein, 2018). This is a problem that has been addressed in financial technology (FinTech) in a variety of ways. As Prado explains, data scientists will often break data up into time intervals. However, this ignores market realities of irregular timing of activities that might be influenced by time of day. To solve for this, data can be broken up by trading volume instead (Prado, 2018, p. 6). This same idea is important to real estate where there might be weekly, monthly, and seasonal changes that need to be accounted for in the analysis methodology. For areas such as commercial asset valuations the challenge becomes even more pronounced, where there is a limited set of assets, and trades happen every 5 to fifteen or more years.

Hiring: Another theme identified by several of the firms interviewed, including Enodo, Foxy AI, Reonomy, and Leverton is that good data scientists are hard to find. According to the CEO of Leverton, companies are probably having challenges applying ML and AI because it takes specialists and it is best to hire a PhD and others

with real experience developing scalable solutions. Real estate technology and other start-ups may have trouble hiring from these specialized talent pools when competing with major companies that have great offerings for their employees (Soman, 2018). Given the potential for ML and AI to be executed poorly, it will be necessary for real estate tech to attract top data scientists and engineers to avoid these pitfalls and create truly revolutionary technologies.

Lack of data: Machine learning methodologies rely on big data, as discussed, but it is also a truism that the more times a model is used, and therefore trained, the better it performs (Finegold, 2018). Real estate is in the early stages of the data gathering and structuring process, so in many cases the uses of machine learning are limited to that task. In the future, that data can be leveraged for analytics tasks, but it could take a long time to get to a point where those applications are truly robust.

2.4 Conclusion

There are many different ways to define artificial intelligence, but machine learning techniques, in particular deep learning, are bringing us incrementally closer to a generalized artificial intelligence. However, we are in the very early stages of using machine learning in real estate, and the areas that could qualify as artificial intelligence are simplistic.

Machine learning can be a very powerful data processing and analytics toolset, with the ability to handle both large and wide datasets of structured and unstructured data. However, it takes a fine-tuned understanding of data science, algorithm parameters, and interpretation to get the right results. Eventually, AI will be able to solve that problem as well, but for now it is important for real estate professionals and tech entrepreneurs to take these issues under consideration.

There appears to be a friction between the promise of machine learning and the reluctance of the real estate industry to adopt a technology they do not understand. Given the explanations provided in this chapter, real estate professionals may be able to make more educated investments in these types of technology and leverage the data they are sitting on today.

In the next chapter we will explore some of the ways that machine learning is having an impact on the industry already. The categories of methodologies, as noted in Exhibit 2, will be explored in the context of the aspects of the real estate industry where they are applied.

Chapter 3 : AI and ML in Commercial Real Estate Today

"All models are wrong, but some are useful."

— George Box, Statistician,
Author of "Empirical Model-Building and Response Surfaces"

This chapter will be an exposé of the current applications of machine learning in real estate technology. The companies applying machine learning today span a wide array of industries so I break the list of firms down into categories of functional areas. In exploring the specifics of the machine learning applications I have selected the primary modes of using machine learning and AI. I will also explore academic research in these areas as applicable.

3.1 Introduction

With this high-level understanding of artificial intelligence, machine learning, and some of the techniques available to data scientists, I was able to conduct interviews and online research to identify 80 real estate technology companies that use machine learning or artificial intelligence.

These companies were broken down by industry area. In addition, through this research nine high-level areas of machine learning applications were identified that are relevant to the industry.

In this chapter, using examples from my research, the various real estate functions where technologies are being applied and developed are outlined followed by an overview of the types of machine learning and AI that are applied most in the context of real estate. More details about the various aspects of machine learning and AI with specific examples from the companies investigated are provided. Additional information from the interviews and literature review for each company is available in Appendix B.

3.2 Real Estate Tech

As discussed in the introduction, there is huge buzz about many sectors of real estate technology, also referred to as prop tech. The real estate industry and the associated services span a wide range of practice areas, which can be broken down in many ways. To help guide discussions around real estate technology for the purposes of this research, companies have been categorized by functional area, each of which is described in Table 3. These categories were created for the purpose of this thesis based on the real estate technology companies researched and a knowledge of the real estate business and how work tends to be siloed. In some cases the categorization is unclear or the company provides multiple services, so the best fit was chosen.

TABLE 3

REAL ESTATE FUNCTION	DESCRIPTION
BROKERAGE AND SALES	Platforms and services for the real estate sales and leasing process
BUILDING OPERATIONS & PROPERTY MANAGEMENT	A range of services and software for managing and operating buildings or tenant spaces
CONSTRUCTION/AE	Construction, design, and engineering industry tools
CROWDFUNDING	Project funding through large groups of people online
DATA & ANALYTICS	General data and analytics providers for real estate
DEVELOPMENT	Platforms or services specific to development investors and underwriters
INCUBATOR	Collaborative programs to help startups succeed
LEGAL/CONTRACTS	Real estate law and contract services
LENDING	Services and tools for the lending/mortgage industry
LOCATION/GEOSPATIAL	Provide services or products related to geography or mapping
SPACE-AS-SERVICE	Non-traditional space leasing, typically with short durations and services from the landlord
VALUATION	Tools, including automated valuation models (AVMs), to evaluate current values and to project future values
OTHER	Other tools that may serve broader industries, such as cyber security, urban farming, insurance, and parking

Real estate technology is receiving a lot of attention from investors and others due to the potential to unlock value in an asset class worth hundreds of trillions globally. According to JLL Spark Co-CEO Yishai Lerner, this is in part due to the attention brought about by the amount of investment capital and large valuations of companies such as WeWork, Airbnb, Compass, and View The Space (VTS). Real estate investors are also taking note and finding that they have the ability to be “king makers”; as first movers to adopt a new technology a large real estate firm can rocket a startup to success, so as investors the real estate companies want in on the action. Simultaneously, or perhaps owing to increased investment, we are seeing that startups in this space have reached a higher caliber since last year

and we are starting to see the first deployments of AI to customers (Lerner, 2018).

However, real estate professionals have long been slow to adopt new technologies. Professor David Geltner observed that as in past technology trends, firms seem reluctant to adopt AI and ML due to inertia, cultural frictions, a desire to not share data, avoidance of accountability, and vested interests in the way things are done today (Geltner, 2018). Part of that reluctance may come from a fear of sharing information that is seen as proprietary, but data is key to machine learning. However, there are shifts seen in the culture of data sharing, as demonstrated by the success of CompStak, a company that trades data for information others have submitted to the platform. They have noted that in large markets

they even get data submitted for the same deals many times over (Lerner, 2018). If real estate professionals understand the potential benefits of the technology, AI and machine learning will be a big component in the growth of the real estate technology sector and, in the next section, some of the impacts that are already beginning will be described.

3.3 Real Estate AI and ML

There are thousands of real estate technology firms out there, including the 2,000+ on the lists that were used as a basis for this research, and that number is growing quickly. However, through the process of identifying companies that employ machine learning or artificial intelligence techniques it became apparent that examples of firms that are employing machine learning and/or artificial intelligence in their systems are rare, and in some cases hard to spot. Exhibit 4 shows the 80 firms that were identified through this research that do use machine learning or artificial intelligence. Not every category is represented here as machine learning and AI are not yet being used in all areas of the industry. There are almost certainly other examples that have not been identified here, and many more to come.

Each of the firms shown uses some form of machine learning, and in a few cases what may be termed artificial intelligence, but the applications vary significantly. For example, some of these firms are using machine learning in advanced data analytics, others have implemented computer vision, and some simply use it to clean data, while still others are combining various techniques. In order to better understand how machine learning is applied in real estate the general themes gathered from research and interviews are examined here. Table 4 includes a representation of the types of machine learning technologies they employ.

These companies are sorted by funding amount within each category, a potential indication of the market interest in the various ML/AI applications and real estate functional areas. As can be seen in Table 4, 65% of the funding is in the brokerage and sales category. Some of these companies require large capital investments for their business model, like Opendoor. However, once single family residential (SFR) focused companies are removed from this list the funding share drops to 29%. Another area that dominates the funding are companies in the building operations and property management realm, which improve operational efficiency and space use efficiency, both of which can be driven by tenant demand rather than landlords in many cases. Other areas are still in early stages of development and have minimal or unknown funding.

THIS PAGE INTENTIONALLY LEFT BLANK

EXHIBIT 4
AI and Machine Learning Real Estate Technology Firms

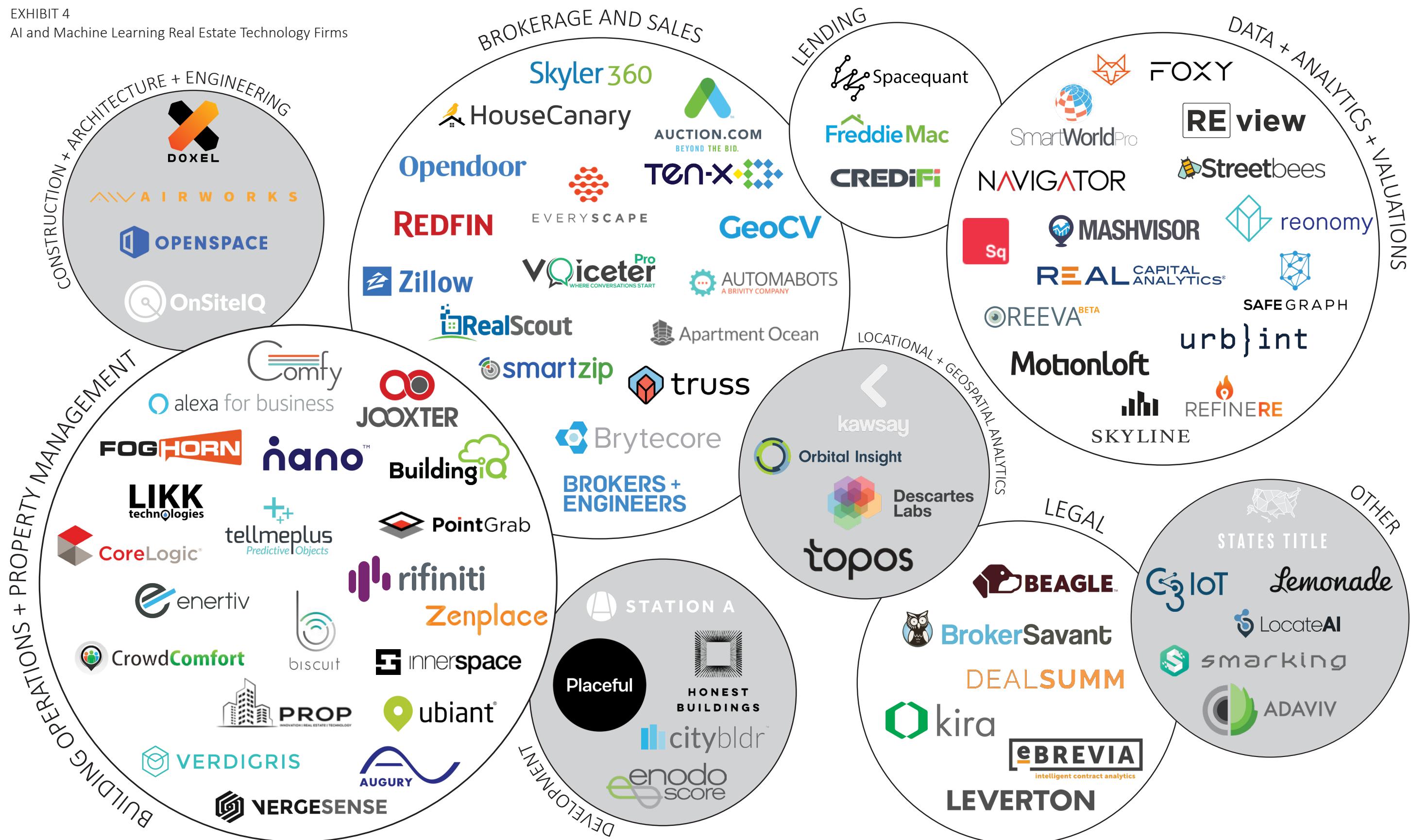
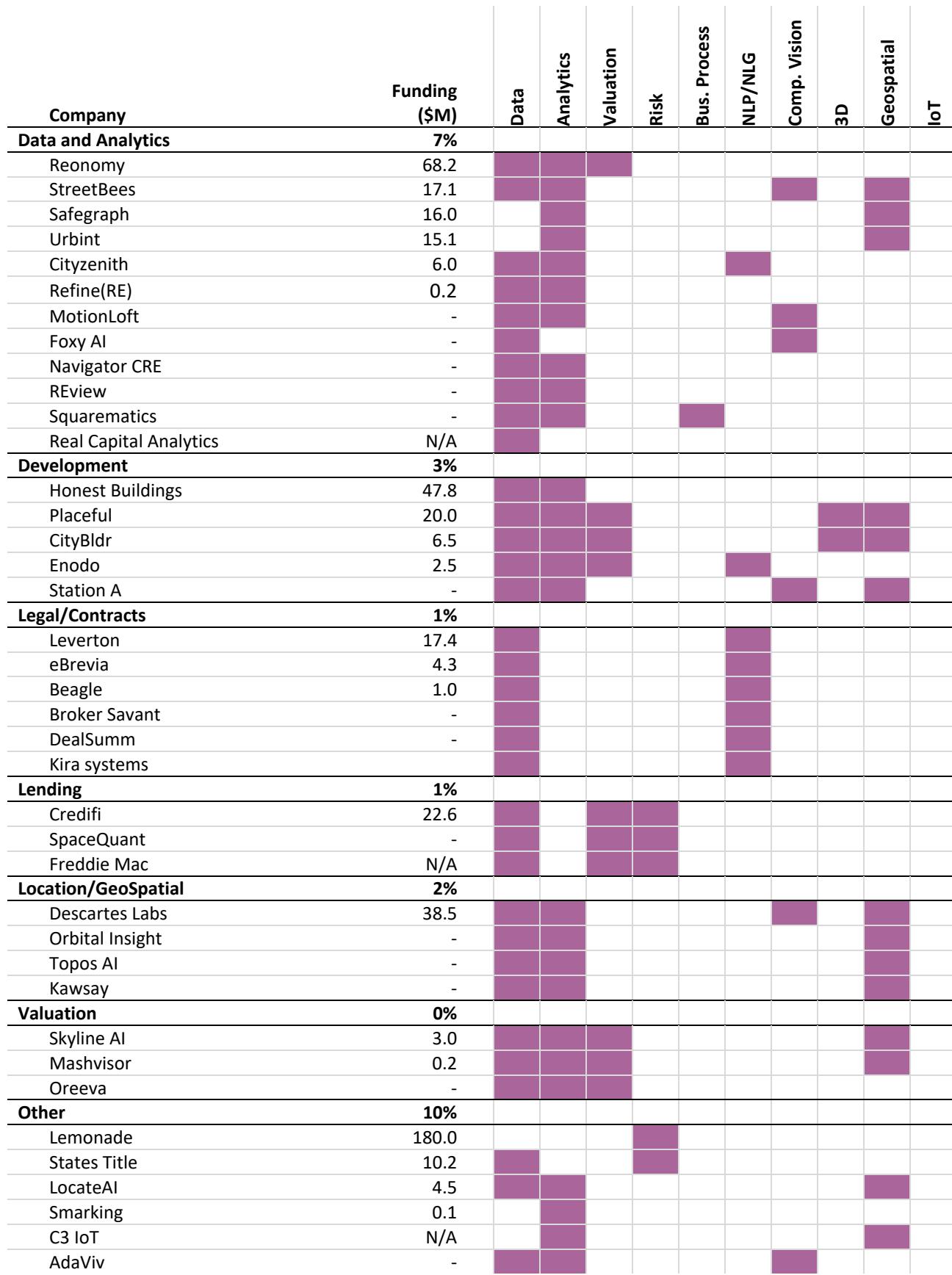


Table 4

Company	Funding (\$M)	Data	Analytics	Valuation	Risk	Bus. Process	NLP/NLG	Comp. Vision	3D	Geospatial	IoT
Brokerage and Sales	63%										
Opendoor	645.0										
Redfin	167.0										
Ten-X/Auction.com	141.8										
Zillow	81.0										
HouseCanary	64.1										
SmartZip	30.6										
Everyscape	20.0										
Truss	9.2										
RealScout	8.4										
GeoCV	3.2										
Brytecore	1.2										
Apartment Ocean	0.2										
Automabots	-										
VoiceTer Pro	-										
Brokers+Engineers	-										
Skyler 360	-										
Building Ops/PM	12%										
Corelogic	50.0										
Foghorn Systems	47.5										
BuildingIQ	26.6										
Augury	26.0										
Verdigris Technologies	23.5										
Comfy	18.6										
PointGrab	12.0										
Enertiv	5.6										
tellmeplus	4.7										
Rifiniti	2.8										
Biscuit	2.0										
CrowdComfort	1.7										
Jooxter	1.1										
InnerSpace Technology	0.2										
VergeSense	0.1										
Likk	-										
Nano Global	-										
Zenplace	-										
Ubiant	-										
PROP	-										
Alexa for Business	N/A										
Construction/AE	0%										
Doxel	4.5										
Openspace	3.7										
Airworks	-										
OnSiteIQ	-										



3.3.1 Data Gathering and Distribution

We start with data, a broad category with many different types of data gathering, creation, and summarization. The real estate tech space is littered with companies trying to make a name as data and analytics providers or finding data sources to leverage for other purposes. Of the 80 companies identified as actively using machine learning or AI, nine fall under the general data and analytics category. However, there is machine learning applied to data in many companies; as seen in Table 4, approximately 73% of the firms have the “data” box checked. This indicates that they use advanced processing techniques to aggregate disparate data sources, clean and structure data, or are generating new data through their machine learning processes.

Everyone uses and gathers data, but these are the companies where machine learning plays a role. The types of data these companies are focused on and the sources for that data vary from application to application. Brendan Wallace, a seasoned technology entrepreneur noted that building a data business is hard, especially when dealing with unstructured data (Wallace & Bruss, 2018). While these firms may not be defined as data businesses, we are seeing that the first step in machine learning for real estate has to be focus on the data, which is a shortcoming of the industry today.

ML can be useful in *feature engineering*, in other words creating new data (Francke, 2018). Firms that are using AI to generate new data for real estate include Street Bees, Foxy AI, MotionLoft, Enertiv, and Urbint. Leverton, a contract data extraction firm, uses *neural networks* to take language extracted from contracts and provide a structured output of the data (Somani, 2018). There are many more companies that are taking existing data sources, aggregating them using ML, and then performing analysis, such as Skyline AI, CoreLogic, Reconomy, Enodo, CrediFi, and SpaceQuant. There are a variety of data sources

that can be used for real estate insights, such as cellular device tracking firms (Thasos, UberMedia), mapping data (open state and city data, Google), demographics and government census data (Esri, Census Bureau), and more.

Data can be a significant source of revenue for the companies that have it. As we have seen with the rise of technology giants like Google and Facebook, their dominance and power is in the data they collect from users. In the example of Google, the company “offers free access to these tools and in return shows you super-targeted advertising, which is how it made \$31.2 billion in revenue in just the first three months of 2018” (Popken, 2018). Several real estate technology companies I spoke with are taking a similar approach to collecting data on user activity and aggregating it to analyze and enhance processes and create better services. These include REview, DealPath, and Truss. Data security and ownership will be paramount to making this work. The hope is that, unlike the consumer world of Google, Facebook, and others, enterprise services have the opportunity to get data ownership right (Somani, 2018).

When data sources are aggregated firms can get powerful insights using many of the other techniques described in this chapter. Data aggregation methods allow companies to collect data from disparate sources and combine them using *record linkage*. Data scraped from the internet can be incorporated into a database of addresses, real estate players, or any other category, and even flagged for important information. Skyline AI has done a lot of this type of work and has “what it claims is the most comprehensive dataset in the industry, drawing on more than 130 sources and analyzing over 10,000 attributes on each data asset for the last 50 years” (Shu, 2018). This includes transactions, location-based data, proprietary databases, mined data, government APIs, stock market data, and more (“Skyline,” n.d.). Their platform then

“compiles all information into a data lake and cross-references everything to find discrepancies and figure out what information is the most accurate” (Shu, 2018). Another example is Locate AI, whose software tracks over 160,000 attributes for each block group and retail tenants to generate analytics to optimize retail location selection (Newman, 2018). This data includes traffic information, sales reports, revenue statistics, location information, and customer demographics and other details found in a variety of sources including UberMedia, a massive mobile data provider (“LocateAI,” n.d.; Newman, 2018). A particular challenge here is instances where there are multiple buildings or assets with the same address. Real Capital Analytics (RCA) is already using this approach for gathering CMBS data, but there could be other applications of this approach in the real estate industry (Goldstein, 2018).

Additionally, HouseCanary “has built a database that includes not only property level info across the country, but also home financing data and interior home characteristics” (Snider & Harris, 2018). Enodo is another leader in data aggregation, and has also identified some more challenging areas for capturing and structuring data, such as lease concessions, and certain amenities, like smart thermostats, which are too rare or new to the market to have good data at this point. Capturing data on views from windows in a building is also complicated with large differentiation in type, quality, and features. (Rutzen, 2018)

One issue inherent to real estate data is that there simply are not enough buildings to make an effective analysis. For example, even with 1,000 comparable leases as data points, it is not enough to get deep insights (Lewis, 2018). Researchers at the MIT Real Estate Innovation Lab are working on an initiative they call the “NYC Wide Data Project” to enable a deep understanding of the impacts of smart buildings, influence of infrastructure, value attributed to design, and much more. Through

this effort they have to contend with the fact that real estate data is often not truly “big,” but it is “wide.” In other words, there aren’t that many rows in the database, but there are more than 4,000 columns that represent a wide array of attributes. The team is assembling a database from multiple sources including RCA, CB Insights, Compstack, Geotel, G Big, and hand-collected data (Fink, 2018).

There is great sensitivity around the types of data that people are willing to share. As analytics becomes more important to every element of real estate, it will be important to figure out what data real estate professionals and related industries would be willing to share, but as of today the collaborative nature is somewhat missing from the industry (Lewis, 2018). When initially conceptualizing the business model for REview, founder Michael Pearce recognized the value of a platform consisting of aggregated anonymized Argus data, with intricate details of asset underwriting, but industry professionals warned him that gathering data from investors would not be that easy (Pearce, 2018). However, it is important to consider the incentives for data sharing and to find opportunities where it makes sense for owners to share their data. As an example, leasing information would be bad for landlords to share because transparency might bring down rents. On the other hand, construction cost data would be beneficial to developers, owners, and investors, and the types of information gathered by companies like Honest Buildings about construction bids and costs could eventually become an important communal resource (Lewis, 2018).

In most cases, there is no advantage from a real estate standpoint not to anonymize data, particularly around people’s movements in and around buildings (Vlaming, 2018). While data is a valuable resource, there are other reasons for companies not to want to “own” user data. Once you own the data you are liable for its use and

protection, and have to consider the implications in various jurisdictions. Most landlords probably have never given any thought to a privacy policy for their tenants. (Vlaming, 2018)

As discussed in Chapter 2, another issue real estate faces is unstructured data. Technology experts who are looking at the real estate sector as an opportunity have to grapple with the inconsistencies in the data that exists today. A simple example of this problem is rental rates; the rent per square foot in an office tower may be recorded as a “gross” number, meaning that it includes operating expenses. However, the building next door may be leased on a “net” or “NNN” basis, with expenses excluded from the number. Furthermore, the space may be measured differently from one building to the next, including or excluding space such as common areas, bathrooms, and mechanical space. There are many different standards of measurement for different purposes (gross, usable, or rentable) and regulated by a variety of entities. This is all part of the problem recognized among several of the people interviewed for this paper; real estate suffers from a lack of structured data on activities and transactions, and it seems at times that the “training data right now is only in people’s heads” (Hong, 2018). With new analytics tools, every interaction with a new software, every sensor output, and every message sent becomes a rich source of data. The hope is that in the right hands, ML and AI will open the door to documenting more data points, and leveraging existing data for actionable insights.

3.3.2 Analytics

For the purposes of this paper, the intention is for this analytics category to cover a broad range of applications where machine learning tools are used for data analysis, whether regression, classification, clustering, or dimensionality reduction problems. Many real estate companies

have lots of data, but are not putting it to use. However, with the growing proliferation of these tools that may change (Wallace & Bruss, 2018; Wang, 2018). As discussed in Chapter 2, “big data” can be an opportunity to find new insights when used well. Approximately 66% of the companies researched use machine learning in some form for analytics.

There are dozens of AI and ML companies that provide data analytics and automation services, and that could be useful to real estate firms or serve as underlying infrastructure for real estate technology businesses. In a list of early-stage enterprise startups, CB Insights identified Element AI, Hyper Anna, Bonsai AI, Peak, and XBrain as young tech firms that provide general service AI and data science tools for enterprises (“9 Early-Stage Enterprise AI Startups To Watch,” 2018). Another general analytics platform is DataRobot, which provides automated machine learning services. The platform supports many kinds of data and will perform *feature engineering* and then test many algorithms to find the one that is most useful for the given purpose (Conway, 2018).

Traditional data and analytics providers include RCA, CoStar, and Ortec Financial. They both aggregate data from various sources and provide reports and indices for real estate markets. All three firms are also working to enhance their analytics platforms with predictive analysis using machine learning (Francke, 2018; Khan & Lupton, 2018; Vlaming, 2018). According to Forbes, “entrants [to real estate tech] are now leveraging machine learning to build upon the early wins in data aggregation by companies like Zillow and CoStar and provide radically more standardized and sophisticated ways to analyze real estate data” (Snider & Harris, 2018). As companies like HouseCanary, Enodo, Locate AI, Navigator CRE, Honest Buildings, CoStar, RCA, and others continue to expand databases and build out artificial intelligence to integrate and analyze it

there are sure to be rapid advancements in data and analytics for real estate.

The firms listed under the data and analytics heading vary in their analytics capabilities and purposes. Reonomy does market analytics and identifies leads for real estate brokers and other related industries. StreetBees automates analytics of crowd-sourced market research. Urbint analyzes urban level data to provide insights that will help shape cities. Chapter 2 gave an introduction to the ways in which machine learning can be applied to data of all kinds for any purpose desired, with the right toolset.

Based on many of the interviews conducted and articles reviewed for this research, it seems that many real estate technology firms are in the data gathering phase of their growth, many of which were mentioned in the previous section. As more users join their platforms and these datasets grow, it will be important to track which companies begin to develop truly sophisticated data analysis methods to leverage that data.

3.3.3 Valuation

Real estate valuation is a key part of real estate in areas such as sales, portfolio management, REIT valuations, tax assessment, and lending. There are many approaches to valuations, whether conducted by appraisers, investors, assessors, or others. Automated valuation models (AVMs) have emerged as a tool to derive valuations more quickly using available data. Seventeen of the companies, 21%, do some form of machine learning based valuation. However, these new tools are not very widespread in use and the results can be controversial. As Bowery co-founder Noah Isaacs pointed out, today AVMs are mostly used for residential properties, and often are not really using machine learning techniques but instead relying on more traditional econometric models (2018). As the data used for valuations gets bigger (and “wider”), machine

learning can be applied to enhance accuracy and speed.

According to the founder of Foxy AI, there are about a dozen companies that produce AVMs and a lot of other companies that re-sell the AVMs, including white labeling and selling as their own (Vomero, 2018). Companies that use machine learning for their models include Redfin, HouseCanary, Oreeva, Opendoor and Enodo. The consumers of AVM results, including financial institutions and large investors, don’t buy just one valuation, they get them from multiple providers and compare the results (Vomero, 2018).

Some sources stated that machine learning AVMs today are directionally accurate and good enough to use when valuing large numbers of properties. However, we are advancing past that use case already, with some companies and researchers able to take a property and perform a full underwriting within minutes, without even having full datasets. We now have enough data to predict what rents would be based on local conditions. As a real estate technology investor, JLL Spark is looking for the companies who are using “deep” data to find investment opportunities (Lerner, 2018).

Zillow has one of the better known valuation models, called the “Zestimate,” which estimates “home values based on 7.5 million statistical and machine learning models that analyze hundreds of data points on each property.” They have brought the median margin of error from 14% when they first launched the product to 5% today using machine learning (“Zillow Prize: Zillow’s Home Value Prediction (Zestimate) | Kaggle,” n.d.). However, the error rates vary by market and the available data (Harney, 2018). It is also possible that the prices being posted on Zillow influence actual sales prices, accounting for some improvement in their predictive powers.

The MIT REIL NYC Wide Data Project discussed in the data section was conceived with the ultimate

goal of better understanding the drivers of asset valuation. For example, the value that can be achieved by bringing co-working space into a building. (Fink, 2018) With the data they are obtaining, the project will likely involve machine learning methods to carry out some of those valuations.

Kok, Koponen, and Martinez-Barbosa have found advantages to applying AVM techniques, including improved accuracy over traditional appraisal methods and the speed of analysis. They applied a regression tree model with random forest and gradient boosting methods to a large dataset of multifamily residential transactions and layered in thousands of variables from a variety of other sources to get predictions with an absolute error of 9%, as compared to an estimated 12% error in traditional manual appraisal methodologies. The results also showed a significant improvement in predictions over traditional hedonic regression models when important variables such as net operating income (NOI) are missing. Another interesting finding is that using larger datasets that included multiple geographic regions actually enhanced the results of predictions of NOI, showing "that the relationship between explanatory variables such as distance to public transport and presence of schools holds across regions." With wider pools and more variables to compare across regions one can end up with better predictions. In testing a multitude of variables, some interesting attributes rose to the top as important predictors of NOI, including the importance of access to green space, music events, and the local mortgage delinquency rates (which served as an indicator of economic condition of the neighborhood). (Kok, Kopopnen, & Martinez-Barbosa, 2017)

Applications of machine learning in real estate appear in other countries, such as China. A paper written by Zhou and Chen of China explore the applications and theory of using neural networks in mass appraisals of real estate in China. They

state that "some scholars combine genetic algorithm[s], geospatial information, support vector machine model[s], [and] particle swarm optimization with artificial neural networks to appraise the real estate" (Zhou, Ji, Chen, & Zhang, n.d.). Based on their findings it appears that in other countries there are already concerted efforts underway to apply machine learning to appraisals and tax assessments. Another paper explores the options for machine learning applications in the pricing of warehouse space in the Beijing area. The researchers concluded that a random forest model produced the best results, with a coefficient of .57 (Ma, Zhang, Ihler, & Pan, 2018). Their conclusion, that the most important factor they studied was the distance to the city center, was unsurprising.

Automated valuation models are gaining in popularity and the machine learning tools applied appear to improve predictive abilities, and also streamline the appraisal and assessment process. However, these efficiencies may not be enough to convince real estate owners and investors to adopt new methods and trust the results; it may be a long time before human verification is no longer deemed necessary.

3.3.4 Risk Assessment

Real estate lending is a data-driven business where banks and other lenders do careful analytics on property values, deal terms, and risk. Insurance is also an area that requires risk assessments. For both applications machine learning is a useful tool to process large data sources and get a better understanding of risk factors that may not have been considered before. Lenders can optimize borrowing levels and rates and insurance companies can balance customer demand and potential costs. About 9% of the companies studied are using machine learning to calculate risk.

Customers of enterprise AI firms, like DataRobot include banks and other lenders who are starting to use machine learning to assess loan risks, though methodologies and uses are not publicly available (“Data Robot AI Experience,” 2018; Goldstein, 2018). Other DataRobot customers, however, provide some insight into how machine learning can be applied. For example, Crest Financial provides no-credit loans up to \$5,000 and used DataRobot to enhance their predictive capabilities to be able to “pinpoint the right type of consumer” and “quickly identify instances where applicants are attempting to obtain financing through fraudulent means, and accurately predict the likelihood of default for applicants.” (“The Making of Data Science Superheroes at Crest Financial,” 2018). Similarly, Zidisha, a non-profit that provides micro loans in some of the world’s poorest countries, uses DataRobot software to improve predictions about loan default and reduce default rates by 5%, opening the door for more loans and giving them a clearer insight into risks (“Directly connecting borrowers and lenders Looking for value in data,” 2018)

Another customer is Freddie Mac, which has used DataRobot to enhance what they call their Loan Advisor Suite, including the Loan Product Advisor; the “automated underwriting system, assesses hundreds of thousands of unique single-family loan files and appraisals monthly.” It also allows for “automated assessments of borrowers without credit scores, immediate collateral representations and warranty relief, and our automated collateral evaluation (ACE), which allows certain loans to be originated without an appraisal” (Ibanez, 2018). These kinds of applications have enormous impacts on borrowers who have more direct access to financial resources and near-instant feedback on loan applications.

The insurance industry is similarly focused on risk. Insurance companies States Title and Lemonade,

are able to quickly evaluate the risk level of each customer using intelligent analytics.

(“Lemonade,” n.d.; “States Title | Crunchbase,” n.d.) Lemonade bills itself as using AI and behavioral economics (it gives all unclaimed money to charity) to optimize renter’s insurance. As the CEO described: “AI displaces brokers and paperwork, and behavioral economics displaces fraud and conflict, all of which results in reduced time, hassle and costs. Aligned interests remove any motivation for not paying claims and checks the inclination to defraud an insurer.” (Fromm, 2017).

With plenty of data points to analyze and significant financial implications, lending and insurance seem to be prime applications for machine learning. Machine learning can quickly reveal prime factors that impact riskiness of a borrower or insurance purchaser, and with the ability to incorporate “wider” data it becomes possible to rely on thousands of different variables that could be risk factors.

3.3.5 Business Processes

As DealPath Director of Marketing Lou Hong pointed out, there are immense amounts of data in real estate, but it is kept in excel spreadsheets and on pieces of paper, not in a format that is easy to use. There is even more data that we are not currently tracking. (Hong, 2018). New software offerings are changing all of that by creating structured environments in which to conduct business practices. These environments can become data sources and can also improve through learning from customer interactions. However, few of the companies, only four were identified in the research for this thesis, use machine learning in augmenting business processes. There are ways of doing this, one of which is known at robotic process automation (RPA), which automates tasks using bots.

There are companies working to automate business practices within and outside of the real estate world, but the ones using advanced computational functionality are limited. Salesforce dominates this market and uses AI with their Einstein tool: "Einstein is a layer of artificial intelligence that delivers predictions and recommendations based on your unique business processes and customer data. Use those insights to automate responses and actions" ("Salesforce Einstein," n.d.). Real estate professionals like brokers and investors could use these types of tools to enhance their processes.

Truss has created a new mode of commercial office leasing processes. The standard process for leasing office space is not transparent, is very time consuming, and leads to less than optimal results for tenants. To fix these problems, educate clients, and save time and effort, Truss offers a service where they have licensed brokers to conduct property tours, but they generally operate using approximately 95% technology, and only 5% human interaction. This is accomplished by gathering information from tenants on their space requirements, including the number of people and growth projections, pin-points on a map of preferred locations, the tenant's industry, space layout preferences, and other qualities they might prefer (for example would you rather pay for more space or nicer amenities?). The service then finds matching properties in the available listings from which the client can then pick out a short-list, set up tours, and go through the lease negotiation process, all on the platform. The software includes an AI chat bot, matching algorithms to find the best properties, and will eventually be able to learn from customer interactions with the software to produce incredibly powerful demand data. (Smith, 2018)

DealPath does not yet use machine learning to analyze data, but they do take an artificial intelligence approach to tasks by automating business processes. Workflows can be put into the

system, and then it uses smart tasks and notifications/reminders to help real estate professionals complete deals. This helps to streamline the workflow and creates consistency in how a business process is run. In addition, they have structured their platform so that searches of documents kept in the system will look at the associations within the workflow to generate more useful results. When enough users are on the platform, there will be opportunities to use analyses of the business processes and outcomes to continue to streamline workflows (Hong, 2018).

Honest Buildings is another example where the business processes are being streamlined through a structured software, but there is not yet any use of machine learning or AI. It is a project management software that allows users to bid projects, track costs, and process contracts and invoices through an online platform. The software is used for capital projects and developments of all sizes (Lewis, 2018).

For many real estate professionals, myself included, processes are ad hoc and tracked through spreadsheets, emails, and paper. Data is stored in siloed systems such as Yardi and Argus. A number of business processes can be vastly improved with cloud based interfaces that at least bring consistency to data collection and processes. However, these platforms must integrate with other key tools, be convenient for the user, and the data must be readily available. Machine learning tools can help streamline certain tasks, but the ultimate goal will be to have true general artificial intelligence that can anticipate moves and make recommendations on the fly when needed. If Salesforce's Einstein or Apple's Siri prove to be accessible and successful tools they could be translated into real estate specific process management systems with truly helpful enhancements to how we conduct business.

3.3.6 Natural Language Processing/Natural Language Generation

Natural language processing provides the ability for computers to “read” text and extract pertinent data. The technology is being implemented for several real estate related uses. The two that are most prominent are contract analysis and chatbots. These technologies are prolific in all kinds of industries, but are particularly interesting case studies for real estate. 20% of the companies researched use natural language processing and natural language generation processes. A few of the real estate applications are discussed here.

Data Scraping: As I discussed with Ofer Goldstein of DataRobot, artificial intelligence tools for natural language processing can be used to take articles and other materials available online and distil relevant information about properties. This is one example of a methodology for structuring data that does not exist in a structured format.

Contracts: One example of natural language processing in real estate is contract analysis and abstracting, including lease abstracting. There are several companies in this space including Leverton, Kira Systems, Beagle, Broker Savant, and eBrevia. These companies are able to deploy machine learning on a robust set of training data, for example the many thousands of existing abstracted leases (Hong, 2018).

The Leverton platform is agnostic to contract type as their software can extract data of all kinds from contracts with adequately sized training sets. Through the use of natural language processing and deep learning to understand sentence structure and syntax, the software achieves two important goals: automating document information capture, and structuring the data so it is usable for analytics. They started with lease abstracting, as it was identified as a clear pain point for many firms. Additional real estate document types that they handle include title documents, deeds, and purchase and sale

agreements. In order to accomplish the contract abstracting process, the machine observes how terms are written in many contracts and learns how to differentiate and record the data in a structured form. The language comprehension element of the software uses natural language processing and neural networks. It can manage many different instances of language use and structuring. For example, one lease could say rent is “\$3 per SF,” another could be written as “\$3 per annum escalating at 1%,” and the next could have a table with a \$3 rent listed under year 1. All of these mean the same thing, and the computer needs to be able to determine the rent in every instance. The accuracy is strong, but still requires human validation for quality assurance, so overall the conservative estimate is that clients get 25% efficiency in the process. These kinds of efficiencies have real implications on time and costs (Leach, 2018; Soman, 2018).

The ability for computers to interpret text enables that text to be structured. With data in this form it becomes more accessible for future analysis. In the example of contracts, the legal profession could get major insights into how certain clauses impact business and legal outcomes.

Chatbots: Chatbots using natural language processing and generation are now a frequent staple of customer service. Most of the tech websites visited during the course of this research had a text bubble pop up in the corner inviting me to ask about the service or software. Some of the easiest applications for ML and AI in real estate are chatbots which can provide customer service, 24/7 tenant service, and other efficiencies (Soman, 2018).

Truss has created a more advanced chatbot tool, using the AI to make significant improvements in the office, retail, and industrial leasing environment for small to mid-sized tenants. Their bot, Vera, collects data on tenant requirements and then guides the tenant through the leasing

process with space recommendations that match tenant requirements, guidance on pricing, and a platform to complete the negotiation of lease terms. All of this has led to a significant reduction in the need for human engagement in the leasing process. For example, one broker in Chicago is managing 130 deals, as compared to a normal brokerage shop where 4 people manage 30 active deals, and it will only get better as the technology improves (Smith, 2018).

However, this does not signal the end of the brokerage business. In an interview Truss co-founder Bobby Goodman noted that:

There definitely is a point where a human gets involved. When we started this, I honestly thought there might be a point where we'd have enough information that a human wouldn't be necessary. It's actually been the exact opposite. The importance of the broker's role hasn't decreased; it's just changed to a more consultative role... The data supports our brokers and our clients in helping them get to a decision faster, but I don't see the need for a human going away. At the end of the day, the people who make the decisions are humans. They have human emotions. They have unique needs. Some clients might need a place for their dog when they go to work. One might need three executive parking spaces. Only a human can take that information and package it in such a way that they're able to negotiate a result for that customer. (PSKF, 2018)

Chatbots and virtual assistants are getting progressively better at holding conversations with humans, even capable of sounding human in certain contexts, like Google Duplex making a reservation at a restaurant (Leviathan & Matias, 2018). Listening to those conversations it is hard

to tell which side is the human and which is the computer.

Document Generation: It's one thing for a machine to read paperwork and extract key data, and another to book an appointment, but a very different task entirely for it to generate business language.

While not fully automated, an example of natural language generation is seen in Bowery Residential's automated appraisal platform. In order to generate an appraisal report in the systems, the appraiser simply uses check boxes to produce a uniform narrative. For example, based on boxes checked, the system might know to generate text to explain that a transitional versus an actual valuation was used, producing a higher value than would otherwise be achieved. Another example might be that a report needs to include a comment on air rights. These types of property aspects are common, and therefore easily repeatable. The software uses natural language generating tools that learn what style of language is needed and then can apply it to the appropriate circumstance. (Isaacs, 2018)

Natural language processing can also be used in scraping all kinds of data from online resources including sentiment data; it is a key component of virtual assistants like OK Google, Apple's Siri, Microsoft's Cortana, Samsung's Bixby, and Amazon's Echo and will be a key component of human-machine interactions in the future. It is possible to imagine myriad applications in machine learning, but the companies that are already pushing this forward are making big advancements in real estate technology.

3.3.7 Computer Vision/Image Processing

One area of machine learning that can be particularly helpful in real estate, a physical asset, is computer vision. Computer vision refers to the interpretation of images and videos, sometimes in

ways that mimic human visual perception and sometimes to pick up on information and gather data in novel ways, like translating images into symbolic data. Popular machine learning tools used in computer vision and image processing include support vector machines, convolutional neural networks, K-Means clustering, and principle component analysis (Bishop, 2006). Approximately 18% of the companies studied use computer vision or image processing techniques for real estate.

A few applications of computer vision include: structure from motion algorithms (constructs a 3D model from a series of overlapping photographs), stereo matching algorithms (builds a 3D model of a building from hundreds of photos), person tracking algorithms, and face detection algorithms (Szeliski, 2010). A YouTube channel called 2 Minute Papers shows dozens of examples of new techniques in image processing, including image stitching and morphing (combine a zebra with an antelope for example), removal of objects from photos, ability for computers to “see” in the dark, and more (“Two Minute Papers,” n.d.). Given the incredibly advanced ways that AI is being used in image processing today, it should be no surprise that real estate tech is taking note and using it to create new data.

Foxy AI is taking a novel approach to image processing for real estate valuation data. They collect data for AVM providers using deep learning techniques where asset quality is determined from property photos. Others, including Amazon, Google vision, and ClariFI have attempted to perform this function with image tagging, but the model ends up being biased based on the features chosen for tagging. They use 30-50 tags for features like hardwood floors and granite countertops, but ignore ugly wallpaper and old, dirty carpet. In contrast, Foxy AI’s model can identify more than 7,000 features. However, the result is a black box, so the users can’t see which features are being considered

when assessing a property. Foxy offers image tagging and room type identification to complement its valuation engine and provide partial insight into what is driving the assessment (Vomero, 2018).

Similarly, Opendoor uses deep learning to identify important value features for single family homes on which they will make offers based on property photos. The qualities considered may include curb appeal, types of kitchen appliances, flooring material, or other important value drivers (Wong, 2017).

MotionLoft uses advanced processing power to analyze video imagery for retailers to better understand foot traffic and how people flow through spaces. They make a device with integrated cameras and processors to count pedestrians, bicyclists, and auto traffic, and to track those movements in and out of retail locations. This data is used by real estate investors to assess the value of retail assets, public agencies for understanding impacts of events and other interventions, and retailers and retail brokers to understand the traffic flows of potential locations. They are also able to track people inside of retail locations to understand movements within a space and assess levels of customer service. The device has been programmed to achieve at least 90% accuracy in every location that is installed, with specialized programming to separate wheelchairs from bicycles and other complex factors. The data is very granular and can provide a lot of insight, and everything is transferred to the software for client review without any images or videos needing to be transferred or post-processed (Wiley, 2018).

Others are using visual data for urban planning and economics studies. MIT Professor Sarah Williams, has directed projects exploring how people use moveable furniture in public spaces. She found that ML for image video processing performed better than the sensors that were

tested in tracking motion and people. (S. (MIT/Envelope C. Williams, 2018). In another example, researchers at Ortec Finance are using Google Street View to see the differences between streets and how they change over time and correlate to other values to get insights into neighborhoods (Francke, 2018).

OnSiteIQ provides a 360 degree video walk-through service to development project managers so they can see the progress as if on-site. They “leverage advanced computer vision and machine learning to automatically map data to the floor plan in less than 24-hours” (“Product | OnSiteIQ,” n.d.). The insights powered by computer vision include the ability to automatically identify OSHA violations, which in turn leads to meaningful project savings (Lerner, 2018).

Another form of computer vision is optical character recognition (OCR). This is an important component of Leverton’s contract extraction process. They built their own OCR software that uses image recognition methodologies to create a more accurate product than existing commercially available options, which have not evolved to process documents that may be in very rough shape. As the suite of documents expands, the software continues to learn new letters, words, and syntax (Somani, 2018).

With the powerful tools being integrated with computer vision technology and the processing power of modern computers, there are many possibilities for how these types of technologies can be used in real estate. The maxim that “a picture is worth a thousand words” is becoming truer from the perspective of real estate investors.

3.3.8 3D Augmentation and Space Planning

In real estate, the buildings themselves are key to understanding the business. AI and ML technologies are enabling a wide range of new

approaches to mapping, designing, and constructing the built world. With machine learning technology, buildings can be quickly drawn and assessed in 3-dimensions. An early example was demonstrated in a 2012 paper where German researchers used support vector machines (SVMs) to analyze coarse 3D models to catalogue building types (Henn, Römer, Gröger, & Plümer, 2012). Currently, at least thirteen companies, or 16% of the list, use machine learning for space planning or 3D analysis.

Today, there are several applications in construction. For example, Doxel sends a robot to automatically traverse a building while it is under construction and then uses AI to assess the imagery collected to track progress and find errors in the construction. To accomplish this, the sensor and image data collected is mapped directly to 3D BIM models (“Doxel AI - Artificial Intelligence for Construction Productivity,” n.d.). Similarly, OpenSpace captures imagery of construction sites through cameras mounted on hard hats, allowing them to track progress and generate maps of the spaces (“OpenSpace,” n.d.). Another construction and design application is produced by AirWorks, an MIT DesignX incubator program alumnus with capabilities to use drone imagery to map buildings and sites into high resolution 2D and 3D drawings. Their software allows a drone to fly around a site and collect survey data with industry-grade accuracy without the need for cumbersome survey techniques. Data collected from the images can also be compared to project drawings (“AirWorks,” n.d.).

3D augmentation and computer vision technology are also being used to map out and virtually stage interior spaces. Using special devices or simply tying into cell phone cameras through an augmented reality interface, companies like GeoCV, Matterport, Google’s ARCore, and many others generate 3D models of spaces to enable virtual tours and interior design. Truss has partnered with Matterport to provide 3D virtual

tours of each property listed on its platform as a way to give the tenant detailed insight into a property before wasting time and energy on tours (Smith, 2018).

Others are contemplating using AI to enable building-level design. While parametric design has been used for a long time to create 3-dimensional designs based on a set of parameters, now AI enables computers to generate designs. An example of this is the platform under development by Placeful. Their idea is to match sites with pre-designed building components to enable housing construction at a lower cost. In a typical residential building, some components are fixed, such as staircases, which have to be a certain dimension, whereas other components are not pre-defined, like room size. Working with a pre-defined kit of components, the software will be able to perform generative design to create the optimal building for each site. This can be viewed as an example of machine learning to solve a *bin packing problem* where the optimal configuration for a set of volumes needs to be determined. The software will also enable the building developers to understand the impacts of trade-offs that will inherently alter the end physical form, for example: should you build out all of the available square feet allowed under zoning if it costs more? Or, should you break a zoning regulation and go to the expense and risk of requesting a variance? These types of second-order effects can be assessed quickly and easily once the impacts are made clear by the model (Fink, 2018).

The product development world is starting to use machine learning for design optimization for all kinds of uses, for example a Swiss firm's neural network produced the most aerodynamic human-powered bike design (Dvorsky, 2018).

As the design and construction industry has moved into the realm of 3D design with BIM models and design and rendering software, the digitization of building data is much closer to the

level necessary to enable complex machine learning interaction. The ability to automatically identify building and site elements makes many processes of design and construction more accessible to a wider audience that may lack expertise and is more efficient at performing complex tasks.

3.3.9 Geospatial Analytics

As we all know, the most important things in real estate are location, location, and location. This is reflected in the uses of advanced geospatial analytics by 35% of the real estate technology companies that use ML/AI. In order to analyze real estate within its physical context it is useful to employ geospatial analysis. WhatIs.com provides this useful explanation for what that means: "Geospatial analysis is the gathering, display, and manipulation of imagery, GPS, satellite photography and historical data, described explicitly in terms of geographic coordinates or implicitly, in terms of a street address, postal code, or forest stand identifier as they are applied to geographic models" ("Definition - What is geospatial analysis?," n.d.).

Real estate technology firms use geospatial analytics in a variety of ways for different purposes. Most of the data, analytics, and valuation companies discussed above use it in some form. There are also companies focused on providing mapping tools, called geographic information systems (GIS) such as Esri's ArcMap, Google Earth, MapBox, and QGIS. Other firms such as Descartes Labs and Orbital Insight are providing information based on satellite imagery, tracking anything from cars in shopping center parking lots to irrigation patterns. Start-ups Kawsay, Urbint, Enodo, and Truss each provide some insight into the types of analysis that can be done using geospatial data and relating it to asset values or other features.

In order to make predictions on all of its data, Enodo categorizes properties by census tract using address information, and then the software uses relevant data, such as demographics, to identify which census tracts in the vicinity are relevant to the subject property. A helpful map visualization of the market areas reveals the precision of the tool in identifying relevant markets and weighting them in the analysis based on statistical similarities, a more helpful methodology than radial analysis, because it finds the most relevant market comparable deals. The clustering algorithms use building characteristics and market performance, looking both at physical nearness and similarities in qualifiers (Rutzen, 2018).

Truss uses geo-spatial data analytics techniques (geostatistical modeling) that they have adapted from mining industries. In mining, a line of gold might be an indication of what you might find along the terrain. This can be translated to a transit line, which provides clues about what might be found nearby. Using these techniques they can understand demand based on neighbors and other local qualities (Smith, 2018).

Truss also takes locational data gathered as tenants pinpoint their preferences and execute leases and using this data they can create a heat map of demand. Property investors can use this insight to track demand movements like a weather pattern, where they've generally observed a traceable path of shifting demand. In the future Truss will layer in a predictive component (Smith, 2018).

Another company that focuses on locational analytics is Topos. They have studied various trends in geospatial terms, such as the paths musicians take when touring the United States, optimal locations for coffee shops, and the differentiators between NYC boroughs ("Topos," n.d.). Topos studies neighborhoods and the relationships between them with a goal of

developing "a holistic understanding of cities through the interconnected lenses of data and artificial intelligence" ("Introducing The Topos Similarity Index and [x] Everywhere," 2018). Their Topos Similarity Index uses a variety of machine learning approaches to gather and transform data, then analyzes the patterns to identify similarities between neighborhoods.

A CBRE subsidiary called Forum Analytics developed a product called ShopoGraphics to help retail tenants determine where to grow. It does this by measuring sales and analyzing the correlation with the other types of retail in a particular location to find the ones that help drive the most business (Stribling, 2018). The software "uses advanced machine learning to categorize over a million retail locations into 37 distinct retail segments based on consumer shopping habits and co-tenancy" ("ShopoGraphics," n.d.).

The MIT Civic Data Design Lab is partnering with Philips on a project called Atlas of Light: "It is an interactive mapping tool in which the user can combine, isolate or cross-reference both quantitative and qualitative datasets to get a clear view of how cities operate. The tool has a potential to create the first comprehensive city dashboard anywhere that digital data exists" (S. Williams, 2018). The platform uses satellite imagery and geo-tagged images from Flickr, Twitter, Instagram, and other sources to enable research on socio-spatial processes and public engagement (S. (MIT/Envelope C. Williams, 2018; S. Williams, 2018)).

Another way that geospatial data is integrated is through software that tracks people using cell phone data. Companies like Google, Apple, UberMedia, and Thasos (an MIT Media Lab spinoff) track people's movements and can sell information about the volume of traffic in various areas and characteristics of the people tracked, like where they go to and come from.

There are numerous sources of geospatial data that can be linked to real estate addresses and lead to important insights about assets, planning areas, cities, and even large regions. With the insights gleaned through geospatial analytics tools combined with the power of AI, we can now understand the complex relationships between many physical aspects of our world: slums and access to resources, subway lines and property value, demographics and building amenities, collocated retail tenants and sales, Instagram posts and neighborhoods, and more. These tools are important for city building, valuation, and a fundamental understanding of how our world interacts.

3.3.10 Internet of Things

Internet of things (IoT) technology is the integration of internet connectivity into physical objects such as refrigerators, thermostats, lighting, sound systems, and pretty much anything else that can be found in or outside of a building. Through these systems, various processes, analytics, and automations can be implemented. Examples of this include tools and devices like Google Home and Amazon Alexa. This is an area that attracts a lot of excitement today, but it is possible we are not even close to realizing the potential of these objects. According to entrepreneur Elie Finegold, one of the biggest applications of ML in the future will be analytics of the data generated by connected “things” which will give us insights into the ways people interact with each other and the world around them (Finegold, 2018). The implications for real estate could be myriad, but we are still in the early stages of applying machine learning and AI to the data sourced from connected objects. Approximately 18% of the companies researched use ML/AI for IoT.

There is a lot of buzz from the residential side around how these tools can transform the way we

perform day-to day tasks, and “smart” features are being built into new developments. For example, luxury prefab and modular home builder Dvele incorporates home automation into all of their designs (“Dvele,” n.d.). These types of tools are starting to be integrated into commercial spaces as well, for example with Alexa for Business, which can schedule conference rooms and alert IT when a printer is down (“Alexa for Business,” n.d.). PROP is a company specifically focused on smart buildings and asset enhancement through technology (“PROP,” n.d.). From a commercial real estate perspective, owners are starting to see the value in using internet connected devices to more intelligently operate buildings through data; now they just have to translate that into deal making (Hong, 2018).

From a building operations perspective, there can be great utility in live monitoring of building systems, from elevators to HVAC systems in order to provide predictive maintenance, rather than preventative or scheduled (Lerner, 2018). Enertiv offers a solution through a platform that captures the most granular equipment level performance data possible and then performs analytics to ensure proper maintenance protocols are met. Their circuit tracking devices connect to circuits that control every aspect of a building system and gets lots of readings from each electrical circuit. They also have devices to monitor vibration, another indicator of equipment problems, and are looking into expanding to other areas. The focus is on quickly highlighting problem areas and ultimately achieving an equipment cost reduction. Using their system, if an issue appears in any equipment, the maintenance team will get an immediate notification with instructions and details of who to contact, manuals and best practices for the equipment, and next steps. With the data gathered on many different systems in different environments they are building a robust

database of equipment lifespans, maintenance requirements, and more (McGill, 2018).

CrowdComfort, however, provides a contrast to the IoT approach, lauding the power of humans as a data source instead. There are many examples of IoT where things that you would expect to be simple and feasible are not. For example, using human networks, CrowdComfort was able to track and identify a trend related to an HVAC system as an improper sequencing, a problem that was missed by the equipment sensors. With buildings constructed in unique configurations and a wide array of installation methods, trying to map a whole building with sensors layers in complexity and makes it hard to work with the system. In contrast, CrowdComfort technology can work anywhere by leveraging building occupants with smart phones to report issues. It seems likely that the human element could add important value to this problem until sensors and IoT becomes more easily scalable and can cover more factors and automate corrections (Graham, 2018).

There is a lot of attention on consumer and enterprise building integration products that fall under the IoT category. Ultimately, the daily interactions that users have with these devices will become data points in themselves, with the goal that buildings and the systems within them learn with the help of machine learning to help the occupants with day-to-day activities, maintain comfort, and respond to requests without specific prompting. The benefits are improved efficiency of buildings and occupants, but there may be a costs to consider in terms of personal privacy and data ownership.

However, the impacts go beyond buildings. With the growth of massive datasets coming from IoT technology we will be able to see patterns in human behavior and interactions with the physical world that we have never seen before. We could potentially leverage this data through

machine learning to solve global problems (Finegold, 2018).

3.4 Conclusion

There are a wide array of use cases for machine learning and artificial intelligence as seen in the examples provided herein. Every company takes a different approach to leveraging the power of these technologies. It seems likely that more firms will emerge in the near term with additional innovations in ML and AI for real estate, including many that are in the data gathering stage today. They will soon be able to capitalize on the boom in the real estate industry's interest in tech.

With 73% of the companies identified as using machine learning involved in data aggregation, cleaning, generation, and integration we can expect major progress in the previously insurmountable problem of unstructured data in the industry. At least 65% of these companies are in early to advanced stages of machine learning analytics and are contending with the potential downfalls of machine learning, but finding new insights that customers need. The range of other machine learning and AI applications across every other category of methodology shows great ingenuity on the part of entrepreneurs in finding ways to reshape the industry.

It is exciting to see the possibilities these firms are exploring, but with thousands of companies trying to solve similar real estate problems there is going to be a “shake-out” of the best in class to allocate customers and funding appropriately – eventually things tend to revert to a handful of oligopolies (Soman, 2018). Real estate investors will want to try to optimize, taking the “best of breed approach” and grow a “tech stack” made up of the best of each stage of data control, but as companies rise to the top they are going to have to figure out how to work together and share the data (Hong, 2018). An approach of collaboration and partnership is something that some

entrepreneurs are touting as key to the future success of their technology businesses (Vlaming, 2018). As real estate investors buy into this new data-driven system and therefore inherently become collaborators as well, it will be interesting to see if there will be a new level of openness in a traditionally siloed and private industry.

There are some hints of things to come discussed by many of these companies. The overall impression from discussions with the real estate technology firms and industry professionals is that we have a long way to go before the machine learning tools reach full potential and we enter the realm of true artificial intelligence in real estate. Those companies that are using machine learning effectively stand to gain a lot, but will have to contend with consumer buy-in in an industry that has proven slow to adapt in the past and that may be reluctant to trust a machine, especially a black box result. Hopefully the insight provided by many of the firms that took time to discuss their technologies and the work and rigor that goes into producing their models will provide some bit of comfort and an appreciation for the non-hype promise of these tools (see Appendix B for more company-specific details).

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4 : Opportunities for Future Applications

"The business plans of the next 10,000 startups are easy to forecast: Take X and add AI."

— Kevin Kelly
Wired Magazine, October 2014

The justification for the use of machine learning and AI in new ways will be explored in terms of the efficiencies and impacts that can be achieved through enhanced analytics. In articles, conferences, and conversations with statistical modeling experts, venture capital funders, and real estate professionals a few ideas rose to the top as potential opportunities for machine learning to be applied.

4.1 Introduction

The 80 or more real estate technology companies that are currently using various forms of machine learning are at the forefront of changing the way we work. Yet, there are areas where it could be deployed in the relatively near term to further enhance the industry further and bring us closer to a general artificial intelligence model real estate.

A key metric in determining the usefulness of a particular technology is its return on investment – is it actually going to create efficiencies and save time and money? Or unlock new opportunities? When I spoke with venture capital fund Fifth Wall the team noted that they are seeing lots of interest from real estate partners in ML and AI, but as an investor Fifth Wall is trying to determine what the real estate industry most wants and what they are willing to pay for. Part of the issue is that we don't know what we are going to find out from proprietary land owner data, so it will be hard to know in advance what value there is in pursuing it (Wallace & Bruss, 2018). With this friction we have to be attentive to the incentives for each party involved in new technologies and think realistically about how they can be implemented.

This chapter will present two case studies for potential applications of machine learning and artificial intelligence for real estate: urban planning, and construction cost estimating.

4.2 Case Study: Data and Analytics, Computer Vision, 3D Augmentation, and Geospatial Analytics for Urban Planning

"We shape our buildings; thereafter they shape us."

- Winston Churchill

Most of the interventions discussed so far have been concerned with asset level, and in a few instances portfolio level, decisions, processes, and tools. These are very useful to real estate investors who need to understand how their buildings are working, what their values are, and how to optimize their leasing. However, concerns around shared spaces and social impacts are going from being relegated to the public domain to being primary concerns of some of the world's biggest developers. By leveraging AI and machine learning real estate investors can ensure that they are creating great places that will attract tenants and buyers because of the intrinsic value of some less tangible assets.

Places like Kendall Square in Cambridge, Massachusetts have become what are commonly referred to as "innovation clusters" where technology firms and their associated businesses, like venture capital and legal services, all agglomerate in a dense area. Today, Kendall Square claims some of the highest rents in the country and continues to attract the best in class biotechnology companies, technology firms, and entrepreneurs. This wasn't always the case, and this kind of success didn't happen by accident. The Brookings Institution released a report on what they call catalytic developments. The MIT-Forest City partnership to develop University Park in Cambridge near Kendall Square, was identified as an example of the capacity for "a dense mixed-use project to create financial returns, generate tax revenue, minimize traffic impacts, and create

affordable housing." Brookings found several characteristics of the projects that achieved this kind of success, including patient capital, employment opportunities, and integrated development with optimal amenities, parking, and retail (Leinberger & Loh, 2018). If these attributes could be repeated and further optimized through the use of technology, developers everywhere could be creating, and profiting from, great places.

Technology entrepreneur Elie Finegold reflected on the role of the real estate business as an experiment in sociology. He sees a structural shift in the way people use space inside and outside of buildings today associated with an accelerating convergence of the development of multiple technologies, including a fascinating opportunity to study new trends. While applications of machine learning are interesting for buildings, it is even more interesting to understand human interactions within and outside of them (Finegold, 2018).

Similarly, in Social Physics, Alex Pentland stated that "we need networked, self-regulating systems that are driven by the needs and preferences of the citizens" (2015, p. 138). With new technologies and analytics tools we can achieve this goal of making socially, economically, and environmentally sustainable cities.

4.2.1 Data and Analytics

As discussed, we are seeing the newfound ability to capture more data as it relates to real estate. This data can be leveraged for many purposes, and at the urban scale can be a powerful tool to shape cities. Two main sources of data could drive this effort. First, the IoT technologies popping up everywhere will be able to provide insights into those structural shifts that are occurring in our interactions with other people and buildings. Sensors and cameras on everything from printers to light poles will track (and hopefully anonymize) our movements. Second, we are carrying "the

most important generator of city data” in our pockets: mobile phones (Pentland, 2015). Some of the companies interviewed for this research are already leveraging mobile data for real estate valuations and other analytics tasks. As mobile phones continue to grow more ubiquitous and are used for more purposes in our daily lives, the data coming from these devices will only become more powerful.

One example of using data and analytics at an urban scale is the much anticipated Sidewalk Toronto project contemplated by Sidewalk Labs, an Alphabet company. As described in the MIT Technology Review,

Sensing and monitoring public activity accurately and frequently will be key. Running autonomous buses on city streets requires knowing when to change lights and other signals to give cyclists and pedestrians priority. Sidewalk Labs says the sensor information would also support long-term planning. The data would fuel a virtual model of Quayside, which urban planners could use to test infrastructure changes quickly, at low cost, and without bothering residents. It could also be stored in a shared repository that entrepreneurs and companies could draw on to make their own products and services for Quayside. (Woyke, 2018)

Alphabet will be aggregating data on public activities, bus routes and usage, and infrastructure, then using analytics to boil that down to actionable information about how cities are used, both at Sidewalk Toronto and more generally.

4.2.2 3D Augmentation and Geospatial Analytics

One major component of the data tracked through mobile devices and other sources is how people move around a city and other geospatial

characteristics of behaviors. Using these tools one can understand idea flows, social ties, transportation flows, and more (Pentland, 2015). Through geospatial analytics we can better understand key features of what drives people and ultimately design better cities. Through the lens of “social physics” we can optimize for the qualities we want in city growth, such as social engagement, creative output, and economic growth (Pentland, 2015).

There are many factors to consider in urban development planning, but machine learning can help sort through the noise to optimize. One way to think about designing cities is through the lens of a game development platform called Improbable. The company develops simulated worlds with incredibly complex interactions of millions of entities (CB Insights, 2018, p. 40). This could be applied to real world scenarios, including real estate development proposals to quickly predict impacts of real estate decisions on economics, traffic patterns, arts and culture, health and wellness, and any other features where data is available. 3D models of new scenarios at an urban scale could also be rapidly generated with user feedback – an augmented public process.

4.2.3 Computer Vision

The technologies to enable broad input on existing urban spaces already exist. An example of geospatial technology combined with computer vision in Chapter 3 was the Atlas of Light project by The MIT Civic Data Design Lab which uses satellite imagery and geo-tagged images from Flickr, Twitter, Instagram, and other sources to enable research on socio-spatial processes and public engagement. As described on the website:

For the first year of the project, the team decided to focus on how lighting intensity varies by various types of socio-economic conditions in Chicago Metropolitan Statistical Area. The tool currently brings

together statistics on demographics, intensity of urban development, and night time light intensity using satellite imagery. In addition to these, the tool also displays quantity and diversity of all the Google Places which is a collection of geo-referenced businesses, institutions, parks, and other points of interest. Lastly, geotagged Instagram posts are used as a proxy for human activity which lets the user to explore the city through the eyes of its users. While an individual geotagged Instagram post might not mean anything, collectively they can be cross-referenced to understand qualitative knowledge about places and interactions (S. Williams, 2018).

Another way to get a better understanding of how people experience and enjoy public spaces within cities would be to employ emotional tracking software. For example, Affectiva has an “emotion AI” that reads people’s facial expressions to “humanize how people and technology interact” (“Affectiva,” n.d.). The insights produced through a human-centered and detailed understanding of what truly makes people happy, energized, and productive could revolutionize the way we think about city building.

4.3.4 Conclusion

As revealed here there are many ways to understand and consider the various impacts an urban environment has on quality of life. Machine learning can be a powerful tool to optimize anything when trained on useful data with the right algorithms. Applied to cities, the outcomes could drive value, create great spaces that people want to be, and help us achieve a sustainable and equitable future.

4.3 Case Study: Business Processes, Data and Analytics, and 3D Augmentation and Space Planning Enhancements for Construction Cost Estimating

In the real estate industry every function touches construction costs in some way. Construction costs drive land values (in that costs will determine residuals), determine asset manager capital budgets, impact feasibility of public and private projects, and determine design decisions. However, the creators and managers of the built world rely on simplifications of industry averages and costly estimating processes to determine the cost of a project or building.

There are existing platforms that are used to track and report on construction costs, but the reports are fairly general, and the data is not always good. Dodge Data & Analytics is one resource for this kind of data, which follows projects from planning through completion. However, researchers have found that it is difficult to rely on self-reported data from building owners collected through a manual process. ML techniques can be used to flag issues in the data and produce insights, for example, the probability of a project starting does not change by product type. However, without enough samples with complete data it is not the ideal source of information (Thompson, 2018).

Other sources are much more granular, such as RS Means, Timberline, and Marshall & Swift, and are used by estimators to build out project cost estimates using available market data. However, these types of tools get very specific in some areas, but rely on personal knowledge to ensure accuracy. They are also cumbersome and time consuming. (Brinser, 2018).

Given the lack of transparency, inconsistency of results, and large datasets that go into making a

cost estimate, it seems like a prime target for advancements in AI and ML. However, the industry will have to resolve a lack of incentives and therefore a reluctance to share data (which is typically viewed as highly proprietary), a lack of urgency from leadership (which views technological advancements as nice to have but unnecessary), and the many complexity factors of building costs.

What are the data, technical, and political requirements to get to the point where we can say “how much will it cost? There’s an app for that!”

4.3.1 Data and Analytics

The data required for construction cost pricing can come from a variety of sources, with several common providers that dominate the market today, but based on new construction technologies there are emerging sources of data. These can be aggregated with other datasets, like building permit applications from the city, weather patterns, locational factors, and more to provide truly meaningful insights that will drive efficiency.

There are different schools of thought for how to approach estimating. Turner encourages its estimators to develop a personal internal knowledgebase, learning from experience and developing intuition about where to look and what questions to ask. Other firms rely more heavily on software and databases of cost data. Typical data sources include spreadsheets and services like Timberline and RS Means. But the data is not enough if people don’t consider the details carefully or factors like economies of scale. One example of the types of considerations that a cost estimator must take into account is, when pricing a floor, it might have the same materials as another floor, but if it happens to have 10 columns in the middle there will be impacts on install processes, timing, and efficiency, and therefore costs (Brinser, 2018).

The cost components of construction projects can be most simply broken down into materials and labor and overhead.

In materials and labor other factors that need to be considered include the quantity and quality of materials, which may vary drastically by building. Determining the material requirements is the first challenge of construction cost estimating and a large source of budget changes over the course of designing a building (Brinser, 2018). As another example, choices like sustainable elements of building designs can have a significant impact. In a working paper, MIT researchers have shown that the costs of green building design for projects in the UK are between 5-7% higher than conventional designs (A. Chegut, Eichholtz, & Kok, 2017).

In some trades materials only make up 10% of the costs, for example drywall, pipes, and HVAC (Brinser, 2018). The remaining major component, labor, is highly dependent on location, market demand, and competition between contractors, and prices can be volatile. Union labor prices are readily available, but actual bids will be dependent on the number of other projects requiring a given trade at certain times and the supply of trained laborers. Contractors have noted that the trades are lacking in skilled workers who can deliver quality work with a high level of productivity, not to mention they are now being challenged with diversity requirements and other hiring challenges. Furthermore, even if new people are hired, a team of relatively inexperienced laborers (the “learner labor market”) is not going to perform at the same level of productivity as the “A+” crew, which will lead to longer construction periods and higher costs. In many cases, time is the biggest factor in cost savings. In 2018 the Boston construction labor market is experiencing major problems with large overhead costs and productivity issues. Most of the major unions in the Boston area are seeing 100% employment and are bringing in “travelers”

who are less familiar with the work. Overall, the trend is a gradual decrease in skilled labor, and the best teams are tired (Brinser, 2018).

There are other factors to consider that will impact costs in addition to materials and labor. To start, for any project it is also important to consider the impacts of time, both in terms of delays and cost escalations over time. Any project delay increases overhead and has implications for income and other downstream effects. In terms of escalations, there are various sources for cost indexes, including Turner Construction and Engineering News Record, which are typically based on a blend of survey data, publicly available materials and labor data, and competitive market conditions, producing a broad understanding of cost trends (Brinser, 2018; ENR, n.d.; "Turner Building Cost Index," 2017). The AGC reports that "the annual rate of increase in the price for new nonresidential building construction jumped from a nearly flat 0.4 percent in 2016 to 3.5 percent in the 12 months ending in September 2017, according to the Bureau of Labor Statistics" (Simonson, 2017). With new tariffs there are more concerns over rising costs, with costs of goods used rising 8.8 percent in one year ("Construction Costs Soar in May as New Tariffs Threaten to Worsen Cost Squeeze, Lead to Delays," 2018). The volatility in pricing caused by tariffs may reflect both uncertainty in the market and opportunistic pricing (Brinser, 2018).

Another consideration at any stage is the difference between pricing a core and shell project and a fit-out. There is also an impact on cost driven by the desire for flexibility within the design and more generic designs, for example when pricing a building that will be built on spec (Brinser, 2018).

Designer and client personalities, preferences, and working styles can also impact the price significantly. For example, jobs that are market-driven are more predictable, versus institutional

projects which tend to have more variety. (Brinser, 2018)

By understanding every aspect of a construction operation, including city authorities, contract management, and staff abilities, these elements of the process can be priced accordingly, while also being optimized. There are many components to cost that a machine could theoretically handle more efficiently and optimally than humans if set up correctly.

A step in that direction is that contractors are also getting more sophisticated with their data collection and application. Some contractors are coding every component they use in their projects so they can then more easily compare projects. This kind of data is going to be key to the ability of estimators to get more precise with their work, but if it is not publicly available we will never have enough data to truly make it "big."

Owners would benefit from construction cost data sharing for their projects as it would ultimately lead to a more competitive construction industry (Lewis, 2018). In the future, bids and actual costs entered into the Honest Buildings platform or other software at a very granular level could be anonymized and aggregated to provide a predictive engine for customers. The data is structured, but the main challenge will be to compare the same costs labeled in different ways (Lewis, 2018). This is the type of task that other machine learning models, like those created by Enodo, Skyline AI, Locate AI and others have resolved successfully with record linkage and other techniques.

Owners are also gaining efficiency through technology. Honest Buildings is a platform that allows users to collect bids online and then track project costs, processes invoices, and manage documents. The software is built in a way that allows users to compare bids and cost information across multiple projects in their portfolio and see

trends and other insights from the data collected through the platform (Lewis, 2018).

When locational information and other factors are layered in, the possibilities are exciting; data can be leveraged to price even uncommon projects, like elevator replacement or new types of green building systems.

4.3.2 3D Space Planning

As discussed, major considerations that impact construction costs are the physical qualities of a building (Lewis, 2018). Contractors are already starting to use more technology and new methods of construction. While automated or robotic building technologies are still in early stages of development, there have been significant shifts towards pre-fabrication of unitized elements or modules in factories, removing some key elements from the chaotic construction site into controlled environments that help to avoid some of the risk factors and creating efficiencies on site (Brinser, 2018). The start-up Placeful is attempting to standardize its building designs and create a modular kit of parts in order to automate and optimize building configuration on a site and ensure predictable costs. This approach will likely not work for the broader development industry, particularly commercial projects in urban areas where everything tends to be custom.

Fortunately, BIM technology provides detailed 3D models that can be automatically broken down into its component parts for analysis (instead of manual take-offs) and other uses.

Other new technologies are helping to bridge the gap between design drawings and actual build-out. Concerns over cost and time to install a floor with numerous columns can be taken into consideration through deep neural networks learning. For example Doxel uses robots and drones that traverses a construction site with lidar and cameras to map out what has been built and how that aligns with the drawings, accurate to

2mm. The technology uses 3D computer vision that is automatically trained to identify building components and is capable of incredibly detailed analysis, such as differentiating pipes. This product has value to identify construction mistakes, but also automated schedule updates with “superhumanly accurate reports” and the gathering of hyper-granular data. Ultimately, the kind of data tracked through this technology, a highly advanced AI, will be able to help contractors make decisions in real time, but also automate the estimation process. Designers will have information about lead times and install process for certain equipment or doors and have that context right at their fingertips. (Tidnam, Lynn, Neff, & Young, 2018).

4.3.3 Business Processes

A Sensera Systems sponsored article postulates about the future day in the life of a project manager on a construction site. He has instant access to information about weather impacts on work on the site that day, cameras allow him to zoom in and see that work is done properly, automated drones are cleared for flight through his app, a BIM model can be overlaid on the completed work to ensure it was done properly, and an automatic notification pops up when there is a conflict with a delivery plan and construction schedules. This is achieved through several key technologies, including BIM, drones, robots, IoT, AR, video and sensor analytics, automated site monitoring (Gaw, 2018).

With all of these technologies in place, and data to leverage, analytics methodologies established, and 3D space planning and augmentation opportunities explored we can bring this to bear on the actual process of cost estimating. Requirements for the construction cost estimating process can be broken into two types requiring different levels of detail and precision.

Type 1: Back of Envelope

First, for real estate developers one of the first stages of any project, generally before acquisition, is a highest and best use analysis of a property, which results in a feasibility study incorporating cost estimates. These analyses can be anything from a “back of envelope” check using market income and cost metrics, taking the average dollars per square foot for the desired building type, to preliminary schematic design cost estimates, working with an architect and contractor. At the conceptual or schematic stage a contractor does not have much design info, and does not know logistics or phasing so the estimator will use their best judgement (Brinser, 2018).

A new application to quickly generate a cost estimate could leverage very large datasets but simplify the output to a general estimate of what a project might cost given a few parameters. A corollary example is an API that was generated automatically by DataRobot to help a hospital estimate the likelihood of patient re-admittance based on a few easily known variables, such as age, reason for admittance, medications, etc., quickly providing a visualization of the probability (Conway, 2018). Applied to a “back of envelope” cost estimate, the machine learning algorithm could work on the backend using trained data to give a price range for a concrete 12-story apartment building with a roof deck and wood floors in the outskirts of the city (or other important variables as identified by the machine). This is an important modification to how real estate developers go about discovering costs and using reports.

Type 2: Detailed Cost Estimate

The second type of cost estimate is a detailed estimate based on concept or design drawings, which incorporate a much greater level of detail based on actual sizing, skins, structures, finishes, etc.

In order to measure buildings and assess material requirements estimators use various software to integrate with design drawings. One example is called On-Screen Takeoff which allows users to outline various building areas in the CAD model to directly get measurements and counts of various elements. Whether using this software or a more manual process, this is an opportunity for the contractor to really learn the job and get a detailed understanding of the complexities involved. (Brinser, 2018)

Once initial costs are obtained through this method, the contractor might verify them with sub-contractors. Unfortunately, they are not the most reliable source of information as they are not incentivized to spend much time on a project that might never happen. There are also checks with engineers and designers to verify that the drawings are being interpreted properly and to flag any major issues. This process can be time-consuming, but some groups, like Turner, have in-house engineering teams that can look at drawings and verify assumptions. This is a move towards the project delivery method known as design-build, which may become popular due to the increased efficiencies of involving contractors early in the design process, and enabling drawings to more easily reflect pricing requirements. (Brinser, 2018)

There are also software options available to estimators, like Assemble, that use building information models (BIM) to automatically create takeoffs, “reading” drawings to measure spaces and count building elements. Since this technology relies on architect drawings, there might be details missed and checks required, especially in first iteration of the designs. The next logical step would be for these models to directly output pricing (Brinser, 2018). As this technology

advances and data analytics can be layered in that process can become more fully automated.

opportunities, assess asset value, and make informed investment decisions.

Architects and engineers will be able to design with price and install time information readily available, avoiding a longer back and forth. One contractor stated that his firm has already been able to cut the cycle of design to estimate to value engineering from eight weeks to a few hours (Tidman, Fischer, Ladha, & Rippingham, 2018). Machine learning and AI could improve this through a more streamlined process with hints and suggestions along the way, while also enabling smooth communication. These tools could also be used to optimally re-allocate the 20% of project costs typically dedicated to rework that have been avoided through early identification of issues.

A platform that uses the principles of data analytics and 3D processes and bring them together would create major efficiencies for estimators. To address the human element of the process, project management software modeled after business process tools like DealPath would streamline communication and enable identification of cost drivers and bottlenecks including staff, politics, and other parts of the design and estimating process (Hong, 2018). Perhaps some of these ideas will come out in fun events, like the AEC Hackathon (<http://aechackathon.com/>).

4.3.4 Conclusion

With the efficiencies of modular fabrication, new data gathering capabilities, and better data aggregation methods it's possible to imagine ML and AI implementation transforming the construction site, and bringing transparency to building owners who want to understand costs. With new data sources and services in their tech stack, developers will be empowered with information that allows them to rapidly evaluate

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5 : Concluding Remarks

"Automation is not our enemy...

Automation can be the ally of our prosperity."

- Lyndon B. Johnson

While real estate technology funding and innovation has come a long way in the past few years, there is still a long way to go. With an improved understanding of the potential applications and opportunities for machine learning and artificial intelligence, real estate investors and technology experts alike can put data to work to improve the real estate business in myriad ways, such as making better decisions, creating efficiencies in workflow, and reducing risk. However, there are limitations to the technologies that need to be considered. There may forever be a time and a place for a simple hedonic regression analysis, a market cap rate comparison, or a human element no matter how far the technology comes. According to Redfin CTO: "We've found that our algorithms work better when we leave a place for a human to be in the loop, and I think that's where the direction needs to go" (Levy, 2018).

As these technologies develop it will be important to carefully examine their use and validity; we cannot assume that a marketing pitch with the terms AI and ML is more than a ploy to leverage the hype. While every firm that claims to use machine learning may be within the definition provided here, there are variations in the capabilities of the data scientists and engineers who are using it and the power of the insights that are produced. A common language about ROI is lacking from this conversation, and until that comes to the forefront it may be hard to get the majority of a slow-to-change industry on board.

However, the tools to enable a broad adoption of machine learning are available if one knows where to look. Automated machine learning platforms are already enabling enterprise customers with data science tools they never could have developed on their own with limited resources. The concerns about avoiding the pitfalls of machine learning are still valid here but with best practices built in, AI for AI can at least open the door to finding some answers within the quickly growing farms / warehouses / lakes of data.

The ultimate goal of a generalized AI that can perform myriad automated tasks and provide recommendations and learn without human intervention in the training element is a long way off. However, there is a lot of room for growth of real estate technology with the tools available today. I look forward to seeing the new applications that engineers develop to make the real estate business more streamlined and the tech smarter.

Technology moves quickly and it may seem difficult to keep up, but if we can see past the excitement of media attention there are plenty of examples of ground-breaking research that is truly opening up insights that never would have been possible before the advent of AI and ML. I expect that the real estate technology landscape will evolve swiftly and it will become ever more challenging to track the companies and research that is changing the way we do business. Perhaps what we need is an AI app for that!

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A - Algorithms

Several common statistical models and machine learning algorithms are described below:

Supervised Learning

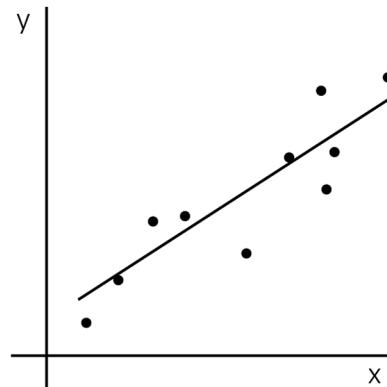
Regression

A traditional statistics methodology, regression is a relatively straightforward approach to data analysis using linear algebra. Regression is used to predict values of target variables based on input variables. Linear regression specifically refers to the use of linear functions to describe the relationship between the inputs and the target. The inputs can be value parameters or non-linear functions that describe parameters, known as basis functions.

In linear and non-linear regression the solution will be found by minimizing the error of the function so that the line fits the data as closely as possible. Examples of the methodologies that can be used are the sum-of-squares error function and the least-squares function. (Bishop, 2006, p. 140)

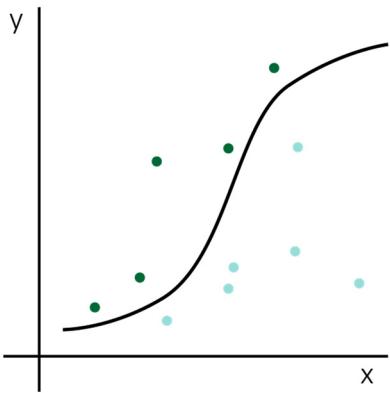
The drawbacks of linear models include the inaccuracy of assuming linear relationships and the difficulty of using many inputs, or multidimensional problems. Nevertheless, these types of functions are important as they are often used as elements of more complex models. (Bishop, 2006, pp. 137–138)

EXHIBIT 5
Linear Regression



A line is found that best fits the data points, expressed in a linear equation.

EXHIBIT 6 Logistic Regression



A curve is found that best splits the data points, expressed in a logistic equation.

Logistic Regression

Similar to linear regression, logistic regression is a relatively simple model that uses a non-linear S-curve function to classify data. It is important to note that this method is used for classification problems, not value prediction problems, the goal of the similarly named linear and non-linear regression problems. (Bishop, 2006, p. 205)

Linear Discriminant Analysis

In linear discriminant analysis, linear classifications and predictions are made in a similar manner to linear regression. However, discriminant functions combine the problem of predicting the probability of an outcome with the decision of how to assign classes or values into one function. This is a simplifying framework, but loses the information about probability of the outcome (Bishop, 2006, pp. 42–43).

Principal Component Analysis (PCA)

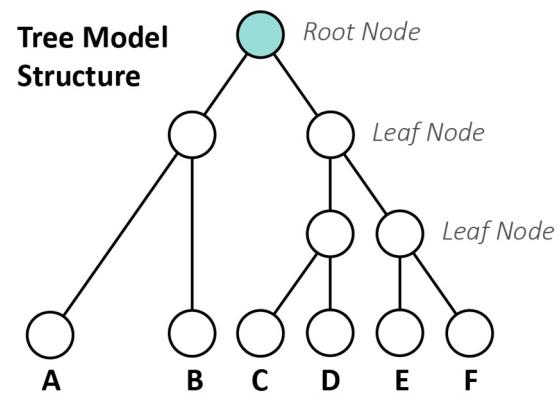
Also known as the Karhunen–Loëve transform, PCA is used to simplify datasets by transforming them into a new set of variables, or the “principal components.” Through the transformation most of the variance can be captured in 1-3 variables. In this method the new variables are derived in such a way as to preserve most of the information about the variance, correlations, or covariance within the data. This can be used to simplify datasets, compress data, extract important features, and visualize data (Bishop, 2006, p. 561; Jolliffe, 2002, p. 1).

Decision Trees

A commonly used supervised machine learning technique is the decision tree. In its simplest form, a decision tree model divides a training dataset along binary paths into subsets repeatedly, generating paths to a prediction dependent on the predictive value of each variable. As a tree model is trained, the structure is determined through a recursive process. All input variables of the training dataset are assessed to find the one that has the highest predictive value, meaning the independent variable with the minimum variance from the dependent variable. Then, the value at which the data is split is determined by finding the point at which each new subset will have the smallest variance in a regression. The dataset is split, becoming new nodes, and the process is repeated, splitting the data again and again. The result is a partitioning of the data, as represented in Exhibit 7, with a separate regression function for each partition to generate the final result. (Bishop, 2006, pp. 664–665; Kok et al., 2017). Tree models can be used for both classification and regression.

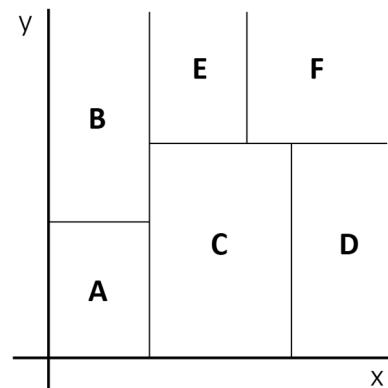
To explain a regression tree using an example, we can consider the problem of predicting the value of real estate assets. We have a dataset of office buildings with continuous variables describing their transacted price (the dependent variable) along with size, proximity to city center, access to transportation, average lease term, age, floors, etc. (the independent variables). Through a regression analysis where the relationship of each independent variable to the dependent variable is compared, we determine that the variable with the greatest impact on value is the proximity to the city center, and that when the data is split at the 1 mile mark we minimize the error of the regression function for each subset. When each new subset of data is assessed we may find that for assets outside of the 1 mile radius the most impactful variable is now building size. However, for the other subset of data, the assets within the

EXHIBIT 7
Decision Tree



The full training dataset is the root node, which is then split into subsets depending on the variable with the highest predictive value. A-F are the resulting categorizations.

Partitioning Representation



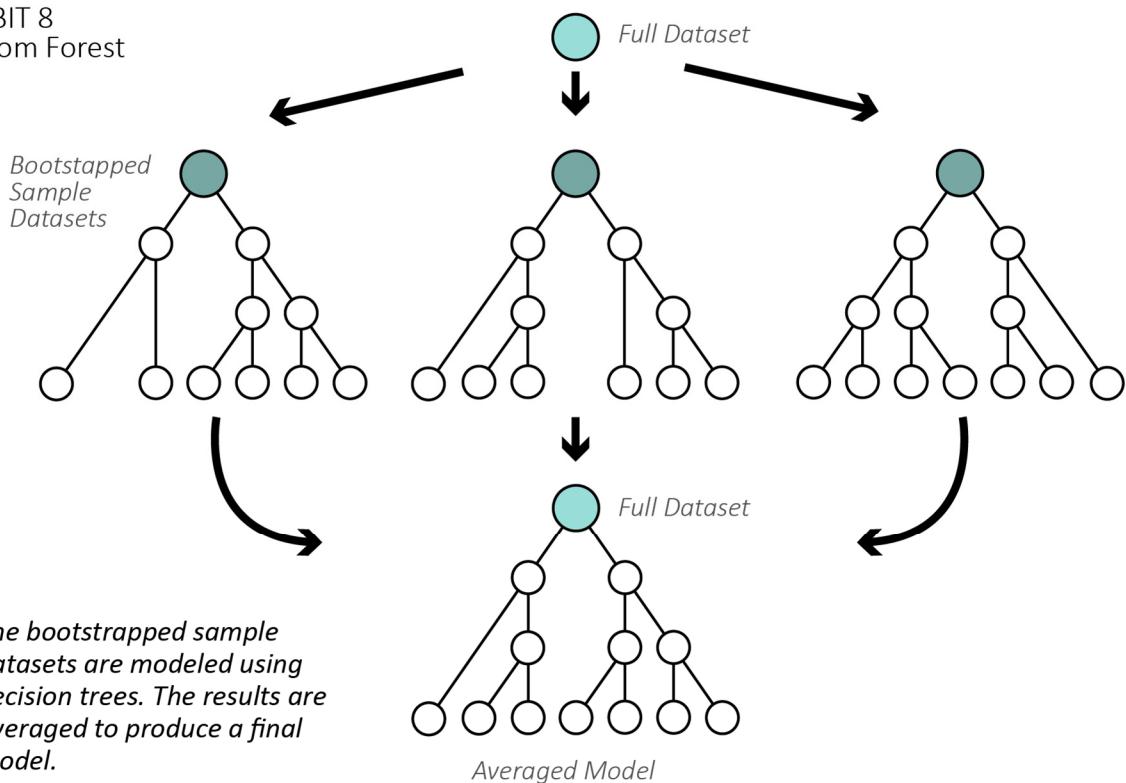
Following the series of splits at each leaf node the data is partitioned into progressively smaller subsets.

1 mile radius, access to transit is more important. These new subsets of data are split once again based on each of these respective variables to achieve predictions. The process repeats until the desired *tree depth* is achieved.

One advantage of tree models is their relative simplicity in terms of explaining and interpreting the outcome (Bishop, 2006, p. 664). To get to a prediction, one simply follows the data point's values down the branches of the tree until reaching the result. An advantage over traditional regression analysis is that dummy variables are not needed, which would typically require an arduous manual process to sort observations into various groups (Kok et al., 2017).

Dangers of using these types of models include over-fitting, with too many branches on the tree, and under-fitting, with too few. Data scientists can assess the ideal depth, or complexity, of the model and “prune” the tree based on various regularization parameters (Bishop, 2006, p. 665). Alternatively, various methodologies can be used to correct for over/under-fitting and other issues associated with tree models, including *random forest* and *gradient/stochastic boosting* (Kok et al., 2017).

EXHIBIT 8
Random Forest



Random Forest

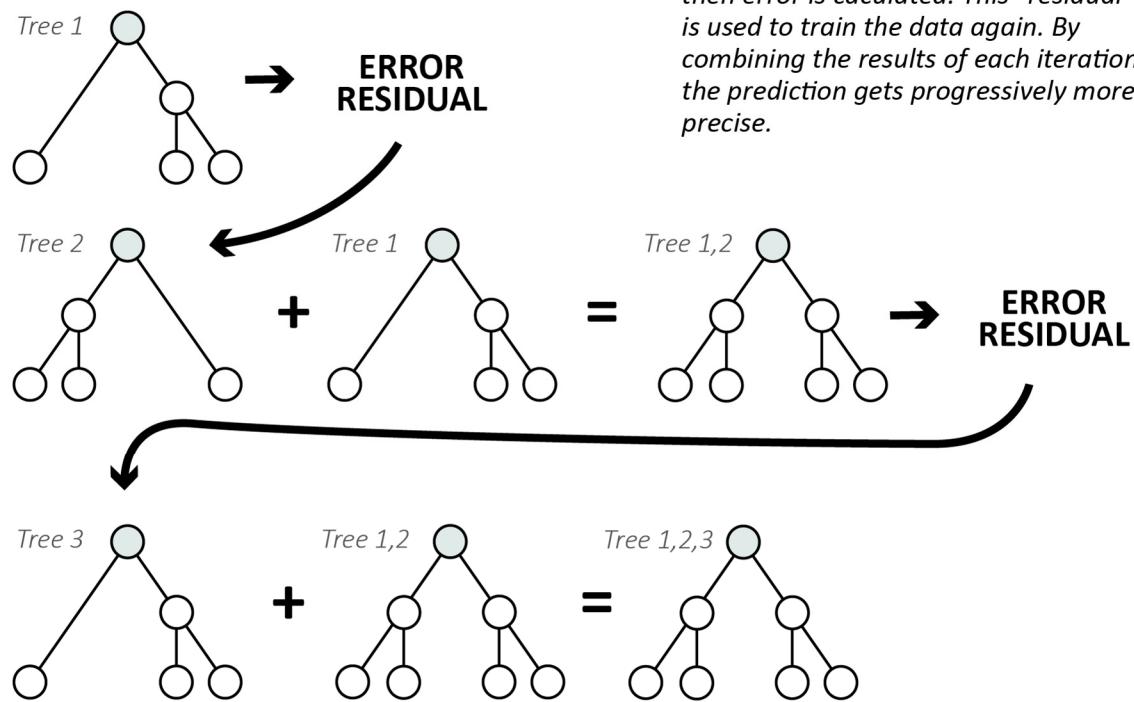
The random forest method is used to correct for over-fitting of regression trees and to allow for testing of more possible relationships in the data by applying the statistical technique of *bootstrapping*. New versions of the dataset are created by drawing observations or data points randomly from the original with replacement, meaning that the same data point can be drawn multiple times. Each bootstrapped sample of the training dataset is used to train its own decision tree model, generating a large number of trees (a forest). When building the decision tree, variables are randomly selected rather than being prioritized by regression analysis. The end result is many tree models produced from many versions of the dataset, which are then averaged to generate the final predictive model. (Kok et al., 2017)

Sequential bootstrapping is another method to improve predictions in tree models and other model types. In this method, observations are assessed for their “uniqueness” in terms of how much an outcome can be attributed to that particular result. The bootstrapped samples are then made up of progressively more unique observations. “Monte Carlo experiments demonstrate that sequential bootstrapping can significantly increase the average uniqueness of samples, hence injecting more information into the model and reducing the “spilled samples” effect” (Prado, 2018, pp. 11–12).

Stochastic Boosting/Gradient Boosting

Stochastic boosting takes a subset of the training data to find a weak signal in the data, and then uses that result as a starting point to create another model on the next subset of data (Kok et al., 2017). A weak model is used to start either a tree or a regression, then a new model is trained using the error residual of the prediction as target variables, adding the results to the previous predictions. The overall idea is that by combining multiple simple models the resulting model can be a stronger model. (Parr & Howard, n.d.).

EXHIBIT 9
Gradient Boosting

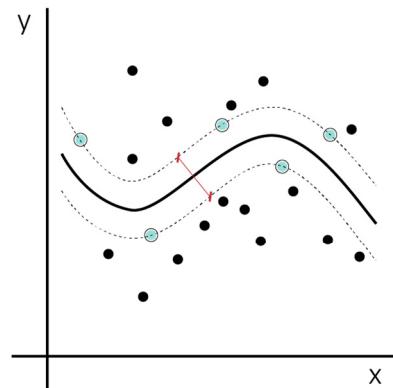


Support Vector Machines

SVMs are a type of supervised learning that is popular for use in classification, regression, and novelty detection problems (support vector clustering methods can be used for unsupervised problems). The idea is to find a *hyperplane* to divide the data such that the margin between the data and the hyperplane is maximized. The points that fall along the edge of that margin are the “support vectors” that shape it. The kernel method can be used in the SVM model to handle non-linear and high-dimensional problems. There is a danger of overfitting using this technique, which can be resolved using what are called “slack variables” that allow for some degree of misclassification. (Bishop, 2006, pp. 325–327, 332)

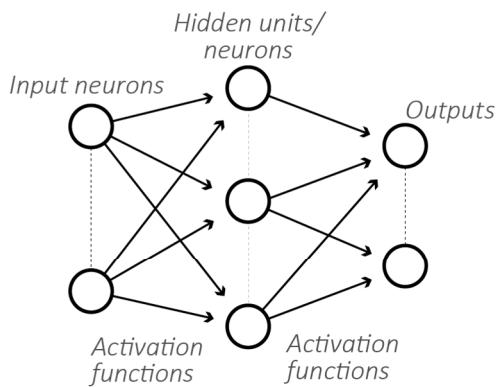
A variation on the SVM is called a relevance vector machine (RVM). Whereas the SVM model is a “decision machine,” RVMs provide probabilistic outputs using a Bayesian technique (i.e. instead of simply returning the most likely classification, these methods return the most likely classification and the probability that it is right). This can be a useful addition to an analysis with improved insights into the results and the ability to combine the model with other models in a more effective way. The results also tend to be more sparse (i.e. using fewer variables) requiring less computational power. (Bishop, 2006, pp. 326, 345).

EXHIBIT 10 Support Vector Machines



The support vector is found by identifying the hyperplane with the largest margin between groups of data points. The margin is represented by the dashed lines and the “support vectors” are shown in aqua.

EXHIBIT 11 Neural Networks



Neural networks consist of layers of hidden units created through activation functions that transform the data to achieve the end result outputs.

Neural Networks

The idea that neural networks actually mimic biological processes for information processing might be inaccurate, but they are a useful model. A neural network model can be described as “a series of functional transformations” of the data. The input variables are transformed using activation functions with resulting quantities being called hidden units. These are then combined to form outputs. “The number of input and outputs units in a neural network is generally determined by the dimensionality of the data set.” The network can have many layers and more complex structures. (Bishop, 2006, pp. 225–230, 256)

To describe this another way:

“A standard neural network (NN) consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment, other neurons get activated through weighted connections from previously active neurons” (Schmidhuber, 2015, p. 86)

Neural networks can be built in a variety of ways, including feed-forward networks with both fully connected layers and convolutional layers, as well as recurrent and recursive networks (Goldberg, 2015). Convolutional neural networks are one type of neural network that are often used in image processing and can be useful for natural language processing (Bishop, 2006; Goldberg, 2015). In convolutional networks the model accounts for the relevance of nearby datapoints using kernel methods, breaking data into subsets that are then processed through a “convolutional layer” (Bishop, 2006, p. 268).

Unsupervised Learning

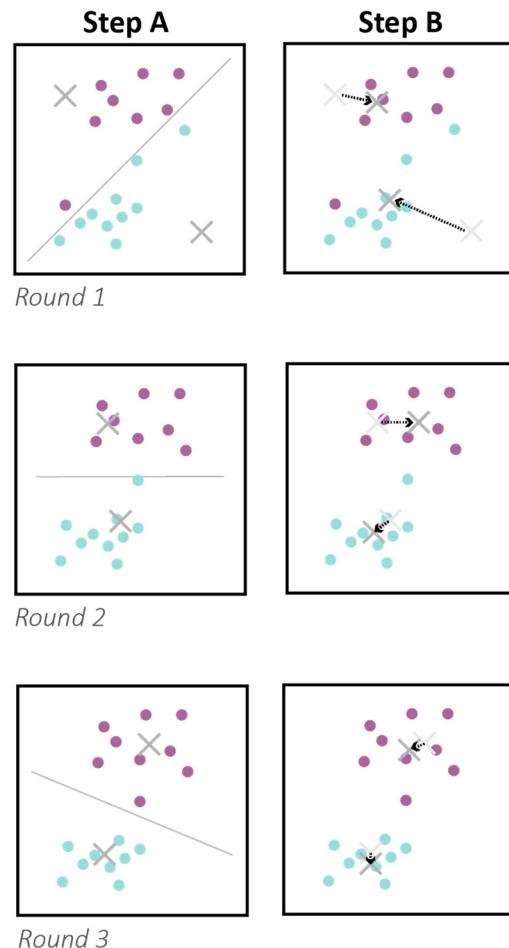
K Means Clustering

When unsupervised classification is required, K Means is a simple way to define clusters of unlabeled data. In this method a set number of categories/clusters is pre-determined (the K parameter). Initially, centroids within the dataset are chosen at random and the data is split into groups based on which centroid each data point is closest to. Then an iterative process follows where within each group the location of a point that minimizes the average distance to all of the data points is identified, and these become the new centroids. Again, the data is split into new groups based on proximity to the new centroids. With each iteration the grouping becomes more rational, and eventually the centroid is fixed because the two processes have converged and the data is split into groups that share the most resemblance. (Bishop, 2006, pp. 424–427)

K-Nearest Neighbors (KNN)

The K Nearest Neighbor approach is used to estimate the density of data within certain parameters or to classify data based on its similarity to other data points within the dataset. A data point that needs to be classified will be compared to a specified number K of nearby data points, and the classification will be chosen based on the classification of the majority of its neighbors. In other words, in this technique the model “remembers” every data point from the training set and assign “to each new test vector the same label as the closest example from the training set.” Of course, this becomes more complex and challenging to execute as dataset dimensionality increases, requiring greater computing power to keep all of the training data needed to specify the model. (Bishop, 2006, pp. 125–126, 291–292)

EXHIBIT 12
K Means Clustering



In each round Step A is to split the data points by closest centroid, followed by Step B where the centroid is moved to the point that minimizes the distance to all the points in the split set. After each move the data is re-sorted in the next round.

Other Concepts

Computer Vision

Computer vision refers to the interpretation of images and videos, sometimes in ways that mimic human visual perception, and sometimes to pick up on information and gather data in novel ways, like translating images into symbolic data.

Computer vision uses a variety of machine learning methodologies, including support vector machines and deep neural networks. When analyzing an image, “nearby pixels are more strongly correlated than more distant pixels. Many of the modern approaches to computer vision exploit this property by extracting local features that depend only on small subregions of the image. Information from such features can then be merged in later stages of processing in order to detect higher-order features” (Bishop, 2006, p. 267).

Natural Language Processing

Natural language processing (NLP) refers to techniques used to allow computers to parse “natural human language” as opposed to the highly structured language computer typically need. There are many approaches to NLP using many of the machine learning techniques already described here. In the past, techniques tended to rely on linear models including SVMs and logistic regression, but more modern techniques employ non-linear neural network models. (Goldberg, 2015)

NLP is used in real estate to read legal documents and power chat bots for a variety of purposes.

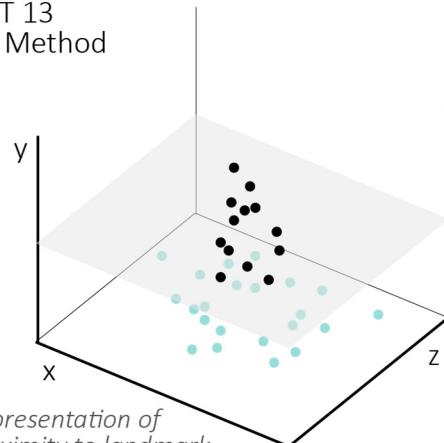
Online Learning

When there are large amounts of data it can be computationally less demanding to process data in a sequential manner, where data points are input one at a time and the model updates as each parameter is input. This can also be done for problems where there is a continuous stream of data being added to the dataset (Bishop, 2006, pp. 143–144). Continually adding new data points to a data set can be an important process for real estate and other models in business, where the model needs to reflect real-time information.

Kernel Methods

In machine learning, advanced methods called kernel methods are used to analyze data without the need to map data into high dimensional space, which can be a very computationally intensive process. One method referred to as the “kernel” function is used to map data based on proximity to landmark points within the data. With multiple landmarks you can quickly achieve non-linear clusters or regressions. The kernel approach is used in popular machine learning techniques, like support vector machines (SVM). (Bishop, 2006, pp. 291–292; “Machine Learning 101 The Kernel Trick,” 2018)

EXHIBIT 13
Kernel Method



Mapped to 2-dimensional space

By mapping the proximity to the landmark onto a 3 dimensional space and splitting the data with a plane the data points are separated.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix B - *Real Estate Technology Company Case Studies*

In the course of researching the numerous real estate technology companies in the machine learning and AI world I had the opportunity to speak with entrepreneurs, technology experts,

Table 5

Company	Real Estate Focus Area	Machine Learning/Artificial Intelligence								
		Data	Analytics	Valuation	Risk	Bus. Process	NLP/NLG	Comp. Vision	3D	Geospatial
Bowery Residential	Valuation									
CityBldr	Development									
CoStar	Data & Analytics									
CrowdComfort	Bldg Ops/PM									
DataRobot	Other									
Dealpath	Brokerage & Sales									
Doxel	Construction									
Enertiv	Bldg Ops/PM									
Endodo	Development									
Envelope City	Development									
Foxy AI	Data & Analytics									
Honest Buildings	Development									
HqO	Bldg Ops/PM									
Lease Pilot	Brokerage & Sales									
Leverton	Legal/Contracts									
LocateAI	Other									
Motionloft	Data & Analytics									
Mynd	Bldg Ops/PM									
Nano	Bldg Ops/PM									
Navigator CRE	Data & Analytics									
Placeful	Development									
Real Capital Analytics	Data & Analytics									
Reconomy	Data & Analytics									
REview	Data & Analytics									
Ten-X/Auction.com	Brokerage & Sales									
Truss	Brokerage & Sales									
Zillow	Brokerage & Sales									

and data scientists. Research and conversations, while not entirely focused on ML/AI, gave some great ideas for areas to pursue where they identified areas for future growth and improved applications for machine learning. Table 5 lists the companies included along with the types of machine learning or AI they use where applicable.

Bowery Residential
Source: Personal Interview,
Noah Isaacs, Co-Founder
Website: boweryres.com
Year Founded: 2016
Funding: \$11.8 million seed funding



Bowery Residential provides third party appraisals by using income and sales approaches. Co-founders Noah Isaacs and John Meadows worked as appraisers before starting Bowery and discovered that they spent lots of time on “busy work,” such as writing formulaic descriptions and gathering asset data. Therefore, the first phase of the startup is mainly focused on automating all that work and enabling them to deliver reports that look better, take less time, and are more cost effective.

With statistics and appraisal experience of their own –in tandem with CTO Cesar Devars, who has an economics and startup background while also being a chief appraiser coming from CBRE with years of experience -- they have together managed to streamline the appraisal process, an \$8 billion market, and continually improve their platform (Loizos, 2018). Currently, their main clients consist of large lending institutions, with a few others including estate planning groups. One of their seed-stage funders is Cushman & Wakefield.

In order to produce these streamlined reports, Bowery aggregates data from 10-15 different websites to populate their report, which involves an automated process and data scrubbing. Later based on observations, the appraiser working on the information uses check boxes to produce a uniform narrative across reports. For example, it might explain that a transitional versus an actual valuation was used, producing a higher value than would otherwise be achieved. Another example

might be that a report needs to include a comment on air rights since these types of property aspects are common and therefore easily repeatable. This process uses natural language generating tools that learn what style of writing is needed and then can apply the appropriate language/syntax to the circumstance.

Phase 2 entails building out a more fully automated appraisal platform based on appraisals produced in the past, but in order to get there the firm will need enough past examples to train their models. Other future plans to leverage the data they are gathering might involve assessing the rental impacts of various building features.

The appraisal process is important to the real estate industry for a variety of reasons. Of course property value is the primary output, but as Isaacs pointed out, institutional lenders and others rely on appraisals as the best source of property level info. Working within the bounds of traditional appraisals but using them as automated data sources can enhance their utility.

Isaacs discussed Bowery’s approach to machine learning and artificial intelligence, pointing out that a general philosophy around building these machines does matter. Humans are subjective, and perhaps we would be better off understanding what the actual truths are, rather than using a machine to approximate how a human would perform a task. An example he provided was a real estate question that an appraiser might face: two widely known truths are that in a city it is better for a building to have more frontage than less frontage, and it’s also better to be on a corner, but which is more important? This is hard for a human to assess, but it is something a computer could more easily calculate. A typical appraisal requires hundreds of these “micro-decisions” but the way it can process that information is drastically different from how a human would go about it.

CityBldr



Source: Personal

Interview, Bryan Copley, CEO and Co-Founder

Website: citybldr.com

Year Founded: 2011

Funding: \$6.5 million

CityBldr is an automated valuation platform that uses lot sizes and zoning to create a 3D massing model of what can fit on a specific site, with unprecedented speed, and then layers in market information to determine its value. The advanced machine learning model can optimize for the “highest and best use” of the land in question and subsequently highlight possible property assemblages and call out the best opportunities for development. It differs from other home valuation platforms because it values assets from the perspective of a developer. Land can be worth a lot more when it comes with the ability to develop multifamily buildings than it would bring in as single family residential development.

Their analysis for buildable space and risk takes into account both “hard” and “soft” constraints. The “hard” characteristics, like lot lines and existing zoning, are modeled in a unique way by CityBldr to establish a value. “Soft” attributes like the politics around development within a community are also considered. For example, housing development may be more politically popular in Seattle than it would be in San Francisco.

Since this is a “black box” technology, there are no users with access to these results. The intention is to raise a fund to invest in the properties identified through this automated process. However, homeowners in the areas covered by the technology can check the value of their homes and get the “CityBldr Value”, which is based on what a developer would pay for the site as

opposed to a home buyer. There are plans to launch a service that will automatically make an offer on homes that qualify for investment based on the company’s criteria, but for the time being CityBldr connects homeowners with partner brokers to market their assets to developers.

This methodology has a clear proof of concept through several case studies where home owners were able to gain significant extra value through property assemblages. These homeowners would have had no access to this knowledge or the developers in the market without the platform. Architects who assessed the sites ultimately came up with plans very similar to those drawn by the software.

Another use of advanced analytics employed by CityBldr relates to buyer analysis. By looking at the history of what buyers have done in the past the software can be used to identify potential buyers for particular assets in the future.

CoStar



Source:

Personal Interview, Philip Khan, Data Scientist

Website: costar.com

Year Founded: 1987

Stock Symbol: NASDAQ:CSGP

CoStar is a leading data provider to the commercial real estate sector. The company has recently started working on new approaches to data analysis incorporating machine learning techniques. Already, their Building Rating System uses ML to predict building value, and has the ability to reveal issues in the data. An interview with a company engineer revealed that there are additional efforts underway, but machine learning analytics is not currently a major focus of the company.

CoStar gets its data by employing 1,600 people to survey the market through phone interviews and emails. They can also leverage online services like apartments.com and loopnet.com for real time listings. Data is collected at the Building level for every commercial property in the US, including multifamily buildings with more than 5 units. The data goes back to the mid-80's, but is considered more reliable and complete after the year 2000. The data collected includes building features and amenities, vacancies, rents, leases, sales, as well as public records.

CoStar has recognized the analytic power they could have with this data if machine learning techniques were applied. For example, this data could be very useful in predicting asset values using advanced automated valuation modeling techniques like those discussed in chapter 3. This would be an expansion of the insight tools suite already offered by the firm.

CrowdComfort



Source: Personal

Interview, Eric Graham, Co-Founder

Website: crowdcomfort.com

Year Founded: 2013

Funding: \$1.7 million seed funding

CrowdComfort is focused on human data – information that can be pulled into the platform from a human (as in using humans) sensor network – which is more sensitive than any device out there. Co-founder Eric Graham seeks to empower people by thinking of them as sensors, and thus making them want to deliver information.

Eric has been involved in energy technology for over 20 years, starting at EnerNOC and working with Kevin Ashton (who coined the term "Internet of Things" while at the MIT Media Lab in the late 90s) in attempting to integrate new energy systems with existing systems. He then went on to MIT, working on applied research in the Fraunhofer Laboratory studying and commercializing sustainable building techniques and solar technologies.

Eric saw an opportunity in that there is a huge emphasis on sensors and IoT, but very little investment in the human side of this equation. The idea for CrowdComfort then emerged: a technology where people can use their mobile devices / smart phones to unlock data about building conditions.

Their self-imposed criteria for the company was that there would not be a need for any wired or battery operated device besides the computers people carry with them constantly. It also had to work on any device in any building. The end solution is an app that building occupants who subscribe to the service can download and use to flag issues at various locations geotagged throughout the building (they currently hold nine

patents on how to geo-locate such data). The occupant experiences are then projected onto floor plans so that building managers can quickly and easily identify and address problem areas. They also create heat maps of what is happening, where it is happening and when it is happening, and will eventually be able to allow customers to generate their own reports such as a trend analysis of how quickly issues are responded to, and where issues are developing. Today this is done in Tableau, a data visualization tool, but in the future they could employ machine learning for more advanced analytics.

They have recently added a feature where QR codes can be used as markers throughout the physical space, providing more visibility for users and making it easy for people to know where to go when using the app.

Their first paying customer was GE, which used an office they had in Bellerica to test ideas. The results were promising; as they found that their existing information flow was 80-90% slower than what they were able to achieve with data provided by users on smart phones. The platform provides a significant improvement in office communication management and safety, productivity, and comfort for building occupants.

Today they have about 50 clients including biotech firms like Amgen, communications firms like Verizon and AT&T, as well as investment companies like Fidelity and Putnam. The platform also works very well for quick service retail franchises.

Eric sees the importance of real estate industry partnerships, and is integrating with CBRE backend systems and other players so that everything is in one place. He explains that Real estate is a big business and it's hard to create a single monolithic platform, but with multiple services integrated the customer can get a better service.

Their use of machine learning is limited today to some natural language processing to translate audio requests / comments into text, but they will be able to do more in the future with larger datasets, such as categorizing something as an AV request. They are also looking at predicting occupant needs.

However, Eric has expressed doubt as to the likelihood of building automation done successfully in the near future, such as having the lighting and temperature set to the right level for each occupant autonomously, due to the immense complexity involved.

Human data is significantly better than the sensor type of approach. There are many examples of IoT where things that you would expect to be simple and feasible are not. For example, using human networks CrowdComfort was able to track and identify an HVAC system problem as an issue of improper sequencing. This engineering error was missed by the equipment sensors. With building in so many different configurations and a wide array of installation methods, trying to map a whole building with layers of sensors adds complexity and makes it hard to work with the system. In contrast, CrowdComfort technology can work anywhere.

Eric also raised the impact of the human experience on better understanding ROI. Without understanding the human element, the tenant is left out of the equation. For example, you may be able to save a lot of money on energy by installing LED's, but then the lighting experience may not be pleasant and you have not really achieved something good. So the question is, if you introduce a new technology to create efficiency, are you really making the impact you want? This can be answered through human sensor networks that allow you to truly visualize the human impact.

DataRobot



DataRobot

Sources: Personal Interviews, Alex Conway, Software Engineer; Ofer Goldstein, Data Scientist; Conference, AI Experience Boston

Website: datarobot.com

Year Founded: 2012

Funding: \$111.4 million

DataRobot is an enterprise level automated machine learning platform. Users can upload data (or add it from a URL or online source), select a target attribute to predict, and the software will then automatically prep the data (including creating holdout sets) and then run all the models that it can apply to that data. A few minutes later the user gets a list of all the models and the outcomes of those models including validation metrics, API exports, lift charts, feature impact reports and more. The machine learning models in the system come from many sources and are automatically customized and combined within the system to achieve optimized results. (Conway, 2018)

DataRobot customer Freddie Mac put it another way:

“What DataRobot provides is the ability to rapidly analyze and experiment with any structured data and then select the most likely algorithms that will give you what you’re trying to predict from it. It doesn’t even need to know anything about the data. It scans and profiles it using unsupervised machine learning, then selects dozens of algorithms to run on the data” (Ibanez, 2018).

I have spoken with several members of the DataRobot team and attended a conference where several companies explained how using the DataRobot platform improved their business

analytics and led to actionable insights. They noted that while data scientists still have their place in assessing the outputs of this software, running data through this automated process saves lots of time, ensures that best practices are being used, and enables a really quick look at whether there is a good machine learning solution for the problem at hand. One customer is Steward Healthcare, a for-profit private healthcare provider. They were able to use DataRobot’s platform to create predictive models to help optimize for length of stay, volume, staff hours and costs, and costs of supplies; all with no in house data science capabilities and expertise. Another example was Demyst Data, a financial services / FinTech provider. Jason Mintz, VP of Product, shared that by using the Data Robot platform and machine learning as a tool for their data analysis they were able to apply an inductive method to their modeling approach, or as he described it, throwing spaghetti at the wall to see what sticks. Without DataRobot’s automation of the algorithm testing process this would not have been possible. (“Data Robot AI Experience,” 2018)

Another type of client DataRobot works with is large banking and lending institutions. According to Goldstein, they are able to provide insights for the consumer side of banking, investment banking, and lending. Lending is an area where DataRobot has been particularly successful, though they were not able to share much detail on this. However, it is helpful to note that institutional types of entities are working to apply these algorithms to their lending strategies. (Goldstein, 2018)

As discussed in terms of the pitfalls of machine learning, there are many challenges to working with time series data. DataRobot has added the capabilities of processing time series to its platform with the ability to automatically detect time parameters and pre-process the data and split it by time intervals (Conway, 2018). Given the

nature of the real estate business, which has a strong emphasis on cash flows and market cycles, time series will be really helpful to real estate research. For example, RCA could use machine learning tools, like those available on the DataRobot platform, to assess same-property transactions that happen at odd intervals if the data could be structured in such a way that the varied time differences could be accounted for (Goldstein, 2018).

DealPath

Source: Personal Interview,
Lou Hong, Director of
Marketing



Website: dealpath.com

Year Founded: 2014

Funding: \$8 million series A

DealPath is an online tool for real estate investment professionals to use throughout the deal sourcing and execution process. Customers are able to assess and evaluate more deals than before with streamlined processes and collaborative tools, while dealing with more complexity by capturing and utilizing data along the way. While in their "Phase 1" they are not using machine learning algorithms, as more deals are done on the platform, data and machine learning will expand the ability to streamline these processes and leverage the data.

DealPath is interested in open access to data and changing its siloed nature in real estate. For example, if you need to get data out of Yardi or Argus and run analyses on it there is no easy way to run queries or use machine learning or other technologies on that data. DealPath believes the user owns the data and should be able to run analyses on it. They recently came out with an open source API, which is attracting interest from other technology companies who want to work with them through that resource. The platform itself has many integrations to import data automatically, from sources such as Google, Esri, among others. In the future, the system could run analytics on deal types and risk profiles, automatically flagging potential issues with a particular property or deal. However, for now they are strictly focused on structuring data.

DealPath collects data on user's workflows and processes, consolidates historical data, eventually using that data to further enhance user insights.

One advantage of using a platform like DealPath is that when evaluating hundreds of deals, one might only put deals that have been executed into the system, but in this format, you are capturing data points on deals that have also been passed on. As the platform grows, it would be helpful to use that history to see trends in the deals that ended up being passed on, seeing how different analysts behave, or seeing if there are corporate models that are more fitting.

DealPath does take an artificial intelligence approach to tasks by automating business processes. Workflows can be put into the system, and then it uses smart tasks and notifications / reminders, thus helping to streamline the workflow and creating consistency in how a business process is run. The DealPath team has seen that often even large firms have no documented or standardized processes for their deals.

In addition, they run searches on documents and data in the system based on associations that are built through the process of tracking a deal. This is another example of an AI principle in streamlining workflows. For instance, if you are working on a restaurant lease deal in NY, you can run a search to get all the past LOI's in your portfolio for any deal done on NY restaurants. This is a way they are using AI principles to streamline workflows.

Doxel

Website: doxel.ai

Funding: \$4.5 million seed funding



In an episode of the a16z podcast, produced by venture capital fund Andreessen Horowitz, the issues faced on construction sites are discussed and some new approaches to resolve them are introduced, including Doxel, as examined by Martin Fischer, professor of civil and environmental engineering at Stanford University; Saurabh Ladha, cofounder and CEO of Doxel; and Christopher Rippingham, who leads technology and innovation leadership for DPR Construction. Doxel uses robots and drones that traverse a construction site employing lidar and cameras to map out what has been built and how that aligns with the drawings, accurate down to 2mm, creating an almost instant feedback loop.

This is key in an industry that is increasingly complex; where a \$100 million project might once have required 100 drawings, today 1,000 blueprints may be seen, or even examples with 15,000 aren't uncommon. In addition, among the many challenges of pricing construction, 20% of project costs are rework costs. Doxel provides a way to reconcile built products in the environment with drawings in order to identify issues earlier in the process, therefore saving time and money.

This technology uses a new kind of 3D computer vision with the advent of Support Vector Machines (SVM), which is able to identify a certain type of object in a specific kind of environment, so it is customized for every object and type of environment. This would not work for something like differentiating installed pipes on a construction site for example. By using this new deep learning model instead, you can "achieve a higher level of abstraction" so the training

happens automatically; since “you don’t have to train specifically for 20 different types of projects”, therefore it is actually learning and bringing knowledge from one site to another. The hardest part in developing this technology was determining how to use 3D computer vision to identify objects in a construction site. A dataset of construction site footage with tagged building elements did not yet exist.

This system allows contractors like DPR Construction to see live progress, automate problem detection, and then quickly make and implement a plan the next day. MEP trades are where you see most conflicts, potential for delays, and equipment interference, these being the areas that are easiest for the sensors and cameras to identify. Schedule updates are automated using “superhumanly accurate reports” from the robot.

The identification of mistakes on the site in real time is key, but there are other advantages to this technology. First, it produces a highly accurate record of what has been properly installed, serving as a reference for the future. A byproduct of this is improved trust amongst trades, owners, and designers. Ultimately, the kind of data tracked through this technology by highly advanced AI, will be able to help contractors make decisions in real time. This opens up possibilities in terms of efficiencies and savings (like the 20% of costs dedicated to rework). Regulatory issues have not been a problem, and ultimately, the team at Doxel would like to have complete trust from inspectors, who could have access to site data on a regular basis, with a third-party verification ensuring that things are being installed properly.

Another major contribution that this technology could lead to is the automation of the estimation process. Every construction project contains millions of objects, therefore knowing hyper-granular data about the details of every object on the site, from design in BIM through construction mapped out by Doxel could be very powerful.

Designers will have information about costs, lead times, and install processes for certain equipment right at their fingertips as they create specifications. The design process will be streamlined, with existing examples where they have been able to cut the cycle from design, through estimate, to value engineering down from 8 weeks to a few hours.

According to Andreessen Horowitz this kind of technological development could be world-changing in other ways;

This technology has broad application to industries of any kind wanting to know what’s going on any physical project of theirs, be it construction, agriculture, shipping, manufacture and many more.
(Dalgaard, 2018)

Not only is the visual power of the machine adding huge value to understanding construction sites, but the data that comes from all of the sites will also be incredibly valuable to the industry, with the ability to automate on many levels. While construction workers may never be replaced by robots, their work can certainly be augmented by them.

Enertiv



Source: Personal

Interview, Connell

McGill, Founder and CEO

Website: enertiv.com

Year Founded: 2009

Funding: \$4.3 million seed funding

The company Enertiv is in the operations management business, specifically focused on improving the efficiency and longevity of building equipment. According to founder and CEO Connell McGill, it is fun to be “lurking in the basement” of the real estate tech space with capabilities that allow them to address concrete problems that have tangible solutions. While ingressing into the field of building operations without much experience, he was initially attracted to this area because there was so little data available on the performance of building equipment. He approached it by building the system from the data up at a time when real estate companies had started to recognize the value of data and technology.

Since its founding, Enertiv has strived to capture the most granular equipment level performance data possible. This began with their circuit-tracking device that mounts next to the main distribution panels to monitor most large systems in the building; with each device capable of monitoring 42 circuits. Data is then collected and delivered over 4G gateways. The software can aggregate multiple circuits that serve the same equipment item. The software then assesses active power / reactive power frequency, and can back into numbers for consumption, demand, power factor, etc. Based on a look at all of the data, the software can find system failure streams. Additional sensor types will be brought into the system as the platform grows. One that has been introduced is vibration monitors on the

equipment itself, which can show motor issues or other failures.

The focus is on quickly highlighting problem areas and ultimately achieving equipment cost reductions. Using their system, if an issue appears in any equipment, the maintenance team will get an immediate notification with instructions and details of who to contact, manuals and best practices for the equipment, and next steps. The standard practice of putting band-aids on systems takes away from how the system was initially designed to perform, so in order to correct for this, Enertiv can take all of the industry best-practices to inform the maintenance team. There is a plethora of information available from the manufacturers, plus there are useful rules-of-thumb in the industry.

The issues and other aspects of equipment maintenance can then be communicated to management directly from the platform with an overview of issues that have come up. This allows immediate insight into ROI, can serve as an accountability tool to show how teams are performing, and with time will provide insight into performance hours and maintenance schedules. Other advantages to building owners and operators include efficiency, for tasks like finding a tripped circuit breaker that once took hours but now instead takes minutes. Or an acquisition requiring in-depth due diligence can be facilitated with this data, rather than relying on a walk-throughs and expense reports. One of the largest RE PE funds interested in this technology consists of an investor looking to leverage this data for acquisitions and asset-level data.

The company is now focused on leveraging the large amounts of data that have been gathered from a variety of equipment in all kinds of environments to be able to recommend preventive maintenance. There are several advantages to preventative maintenance including a potential for 75% reduction in

equipment costs and the extension of equipment lifetime by about 40% on average. A study produced by JLL showed a 500-600% ROI on equipment maintenance practices. All of this adds up to much more than what you might save on energy costs with marginal improvements to equipment run times. From a leasing perspective, it is important to look at systems before they fail and cause problems for tenants.

To start out, Enertiv has targeted un-or-under-instrumented owner-operated office and multi-family portfolios, instead of new class A buildings, which tend to have lower maintenance and equipment costs. One customer is a leading industrial operator, and having started with almost no data on their equipment, they can now look at hundreds of thousands of circuits throughout their portfolio thus optimizing their equipment maintenance and purchasing plans. They are also targeting select service hotels which have lots of AC units and other equipment that is dispersed and hard to manage.

By using ML to assess all of this data, they are getting progressively better at identifying the problem area. With more data they can improve predictions and are currently finding three or more new insight types every month using ML algorithms. They are also asking the users on all sent notifications whether the information was helpful, allowing the computer to learn through human inputs and interactions with the application as well. Eventually, owners will be able to get incredibly accurate equipment performance ratings, understand which equipment performs best in each environment, and help prioritize investments in capital projects.

As they continue to grow and develop their capabilities, Enertiv will seek to become the go-to one-stop-shop operations data aggregator. For the future they are looking into other mechanisms, meters, and sensors, like condensation, and more.

Enodo



Source: Personal

Interview, Marc A.

Rutzen, Co-Founder and CEO

Website: enodoinc.com

Year Founded: 2016

Funding: \$2.5 million seed funding

Enodo is an automated multifamily residential underwriting platform that allows real estate developers and landlords to assess building value and to get real-time information on building amenity values, based on relevant comparable sales or lease rates in local markets. The analysis can be used for existing buildings, re-developments (value add), or ground-up projects.

In order to create their analysis, Enodo pulls in data from a wide array of sources. As noted on the Enodo website, “there is no shortage of data when it comes to commercial real estate, from open sources to paid data providers, there is a wealth of information in the marketplace. The difficulty is how to bring all this data together and try to make sense out of it to help make better business decisions” (Rutzen, 2018). The data sources used by the Enodo platform include property management software like Yardi, direct rent roll uploads from users, direct connections to over 37,600 property websites, 11 different listing sites, city open data sources, Esri (for demographics), and Google maps (for location information). Given the diverse data sources, Enodo puts 90% of its efforts into cleaning the data, removing outliers, and combining diverse datasets into a structured platform. To accomplish this, they use a variety of machine learning algorithms, such as sophisticated outlier detection algorithms. For example, if leases are those of shorter term they can skew rents, or if they sit on the market for too long they may not really be representative of the market.

There are many features that define a property, and Enodo is working to capture as many of them as possible; however, there are some areas that are more challenging than others. For instance, it is hard to get data for lease concessions, and there is not yet a way to quickly train and structure the data, including items such as waived application fees. In addition, some amenities are far easier to get data for than others, such as in the case of smart thermostats which are too new on the market to have good data on them at this point. Views are also complicated with large differentiation in types, quality, etc.

Taking data from listings, property websites, and user uploaded information requires processing to identify the qualities and amenities of a space. Part of this process is something they call the “Booleanizer,” which identifies similar amenities that have been pulled out of language from space listing text. There is also a sequential logic built in, such as: if there is a concierge, then the building has a lobby. Additionally, there are methods called proximity negation terms that come into play, as in: if the word “spa” occurs by itself, there is a spa in the building, but if it appears with certain other words then it is negated.

In order to make predictions on all the available data, properties are categorized by census tract using address information, and then the software employs relevant data, such as demographics, to identify which tracts in the vicinity are relevant to the subject property. The platform will then use computer analytics and machine learning to find prices and demand for different unit types within the market and assess the attractiveness of the property being analyzed.

A helpful map visualization of the market areas reveals the precision of the tool in identifying relevant offers and weighting them in the examination based on statistical similarities, a more helpful methodology than radial analysis because it finds the most relevant market comps.

The clustering algorithms use building characteristics and market performance, looking both at physical proximity and similarities in qualifiers. The machine also picks a data set that will be large enough to generate results for the particular attributes that are relevant to the subject property. This is accomplished using K-NN clustering and then overlaying custom code.

Other types of algorithms are used in the software as well, like natural language processing which can get the rent roll and operating statement from a user uploaded file, including PDFs, or property websites. Outlier removal methods are used to clean the data by comparing samples at market level, state level, and census tract level. Part of the process of automating this feature involved plotting data on maps and looking for patterns that would be a first step in the analysis process that served as hints in writing the algorithms. Despite all these advances, there is still some manual / human involvement today. If issues in the data are flagged, a data scientist can evaluate the validity of the data and determine whether it can be incorporated into the platform or not. For example, if data comes in indicating that a studio has 2 bathrooms, it doesn't make logical sense, and it will consequently be removed.

Eventually, Enodo plans to expand their platform to all property types, starting with single family residences (SFR), and potentially hotels, which are similar to multifamily in the amount of information available online and the types of amenities.

Envelope City

Source: Personal Interview,
Sarah Williams, Co-Founder

Website: envelope.city

Year Founded: 2015

Funding: \$4 million

Envelope provides spatial zoning interpretation to help investors and developers understand development opportunities. The founders felt that zoning should be a system that can be digitized, but in today's world it is unfortunately text-based. To solve this issue, they took New York City's zoning code and created a tool where the zoning envelope, including height limits, setbacks, and floor area allowances are readily available in digitized 3D form.

There is a substantial value in this digitized record of the zoning code, where in many ways the product is the data. To get the zoning code into this format the team used two separate processes: the first consists of hand digitizing a database of combined codes and interpreting what they do / mean, and creating a look-up database. They were able to use city resources as a baseline, but sadly some of the city geometry data was inaccurate and required correction. The second process was one dealing with visualizing all this data into 3D, which was also a technology they developed in-house. Now that this process has been completed it makes it repeatable for other cities, especially those with open data sources, like open GIS tools. The team is highly confident that this is a very scalable approach and will be starting in larger urban markets with structured zoning (and not places like Boston, for example, where zoning is less structured and has more complicated permitting requirements).

The tool started as a service where users could find a property they were interested in and getting a quick assessment of building envelope



options. In the past few months they launched an additional service where clients can use their speculative search engine to identify opportunities based on site requirements.

Envelope City's founders are experienced with machine learning techniques but do not use it at this point, the dataset size does not require that kind of processing. However, there are a few potential future uses. One example would be trend analysis, which in addition to speculative search, does not need ML because people's requirements are fairly straight-forward. Nevertheless, machine learning could help when comparing sites across different cities, besides overlaying other data sources which would be another step requiring ML processes.

Foxy AI

Source: Personal Interview,
Vincent Vomero, Founder

Website: FoxyAI.com

Year Founded: 2017



Foxy AI is a data provider analyzing property image data for companies that are creating AVMs, to both investors and analytics service providers. They use computer vision and deep learning for property images and take unstructured data to create structured information.

There are about a dozen companies that produce AVMs, and then several other companies that re-sell the AVMs (including white labeling and selling as their own). The consumers of those AVM results, including financial institutions and large investors, don't buy just one valuation, but instead get it from multiple providers and ultimately compare the data. While founder Vin Vomero explains that they are not trying to compete with AVM providers, they have built their own decision tree model based on their data and achieved good results.

The value of Foxy is in providing more accurate information for properties situated at the "extremes" of the market spectrum, whether very old and run down, or brand new and high-end. Traditionally, AVM providers have struggled to account for the quality and condition of the property, relying on size, number of rooms, and other commonly tracked characteristics. As a result, the models work well enough for average properties, but they do not work as successfully for outliers. As automated valuation models become more popular, and even more useful in lower end markets, where an appraisal is only required for any sale above \$250K, properties below that threshold can simply utilize the automated model.

To produce data, Foxy AI found success by requiring property photos obtained from customers, as opposed to others which have attempted to use image tagging services, such as the ones offered through Amazon, Google vision, and ClariFI Software, a method that does not work for valuations. When doing image tagging, the model will be biased based on the features chosen for tagging. For example, most people are tagging features like hardwood floors and granite countertops, but ignore ugly wallpaper and old, dirty carpet, and as such, 30-50 tags is not enough to make an informed valuation, whereas in contrast, Foxy AI uses a deep learning model that can identify more than 7,000 features. However, the result is a black box, so the users can't see which features are being considered when assessing an image.

In order to improve the software and manage this black box issue, Foxy AI has a separate identification platform to identify room types, and will build this out to allow for a classification network to tag features, which can then be used to give some interpretability to the model. The Foxy AI product offerings are:

- House2vec – the deep learning model
- Room Type Labeling
- Object, finishes & scenes detection – detects granite but also water stains, bad wall-paper, etc.
- Comparable properties

Interpreting deep learning models is a big research area, as demonstrated by the IBM partnership with MIT focused on solutions research.

Vin first got interested in AI a year before conceptualizing the company in 2016 while working with his AI expert, he started building out the models in late 2017. Vin noted that it is hard to find people with expertise and real-world experience in computer vision, but fortunately his

first employee had lots of experience, having spent 10 years working in computer vision.

Honest Buildings

Source: Personal Interview,
Geoffrey Lewis, Head of Product

Website: honestbuildings.com



Year Founded: 2012

Funding: \$47.8 million

Honest Buildings is a project management platform that allows users to bid on projects, track costs, and process contracts and invoices through an online platform. The software is used for capital and development projects of all sizes.

While the use of machine learning analytics is limited today, Honest Buildings was built in a way that allows users to compare bids and other information across multiple projects in their portfolio. As these comparisons are constructed from within each customer's database, they can assess the opportunities to leverage that information for predictive analytics and tools that use anonymized user data to show market trends and projections. However, the first step is to make a usable product and begin to encourage consistent data entry within the industry, while respecting customer needs.

In our conversation, Chief Product Officer Geoff Lewis reflected on the limitations of the data available in real estate today and the reluctance of real estate professionals to share that information. However, real estate investors would better benefit from construction data sharing to drive efficient markets and, therefore, lower prices. Nonetheless, even if this data did become available, it may not be useful for certain project types or projects of a particular scale simply due to the limited number of buildings that are constructed. For example, you may benefit from many comparable office fit-out projects, which happen frequently, but there are not very many comps for building elevators in high-rises.

There are also other challenges to analytics in construction cost pricing, including physical differences in building design, location specifics (how the space is accessed), and other factors. There will always be a point at which estimators are required to give an accurate bid on a particular design.

Honest Buildings is creating a popular platform with a growing set of users. Eventually the data collected through users may serve as an important tool for in depth analytics of both project management and cost projections.

HqO

Source: Personal Interview, Chase Garbarino



Website: hqo.co

Year Founded: 2017

Funding: \$4 million

HqO, or Headquarters Optimized, is a tenant engagement service and platform that provides a central application for all tenant concerns, needs, and amenities. Founder Chase Garbarino, looks at real estate technology in today's early stages and imagines the corollary scenario occurring parallel to what has happened in the marketing world; where it evolved from simple services like email marketing (which became a huge business), to services that tackle marketing from several angles and where email campaigns are just but a small part of the business as a whole. The software they are building is flexible and can be shaped as it learns. The company endeavors to create value for landlords while making tenants' days more efficient.

HqO looks at the data it collects as belonging to the client, such as the building owner; in part due to considerations of data ownership. There are externalities not often considered in relation to data ownership; where it requires a privacy policy in consideration of liability and jurisdictional differences.

HqO is looking into integrations with other technologies, allowing their platform to aggregate all services within the building. Today they are integrated with transit systems, shuttle services, building access security systems (HID), and activity programming uses.

Future applications of the types of data collected through the application could include insights into tenant activity and how best to program buildings. By knowing who is taking advantage of deals or

activities, tenant demands can be better understood and inspire future amenities and activation plans. It could also provide insight into transit and shuttle use and their respective patterns. Eventually this can be accomplished using machine learning, but capability will need an estimated 3-5 years due to the large amount of data needed to truly leverage those tools. Ultimately, this could become a recommendation engine that prescribes optimal amenities and other building features to customers.

LeasePilot



LeasePilot®

Source: Personal

Interviews, Gabriel

Safar, Co-Founder and CEO and Itzik Spitzen, CTO

Website: leasepilot.co

Year Founded: 2015

Funding: \$1.5 million seed funding

When Gabriel Safar started LeasePilot, he envisioned a platform that could automate lease negotiation and writing, thus predicting where the terms would end up. Today, the platform provides a way to take an unstructured and cumbersome process and creating a streamlined solution, such as taking word documents for standard leases provided by customers and converting them into structured XML files. It continues to look like a document to the user making edits, but concepts within the lease are tagged to the appropriate parameters, like rents and start dates but also more complicated factors for example whether the lease automatically subordinates and how the options are structured. In the end, every business term becomes sortable, filterable, and available for analytics. By starting with a lease in this manner, there is never any unstructured data to worry about improperly abstracting. LeasePilot allows you to institutionalize the lease process and make the architecture of the lease transparent to all parties involved. By creating a lease in this fashion, they are fixing a broken system while eliminating the need for lease abstracting and natural language processing to structure the data.

Once the terms are in the system the customer can easily make modifications to the lease since the software now has the ability to provide “guardrails” for this process, where the user can set up parameters for the lease terms. Eventually this will remove or limit the need for lawyers in the process, especially on smaller and repetitive lease deals. This creates huge savings for

landlords and tenants who spend lots of money on leasing fees any time attorneys are involved, and who are frequently hurt by delays due to extended negotiations.

This is done through a manual tagging process that differentiates them from natural language processing-based leases, thus abstracting providers like Leverton. However, when this is accomplished for one lease, this is forever taken care of, therefore avoiding the stage of the lease negotiation process where business points are taken from the LOI and turned analogue in the legal document, obscuring the structure of the data.

The LeasePilot technology is generally applicable to any kind of agreement; since there are other types of agreements and contracts that behave very similarly to leases, such as, loans, M&A agreements, insurance policies, and more. In essence, every document where one party “owns” the document and the options available for standardization can be applied would work under their model.

The software is built to speak the language of attorneys, generally being the parties who are making alterations to legal terms of agreements and formalizing business terms. Lawyers tend to be tough customers and slow to adopt technology, but with attention to the UI added to the great efficiencies achieved through the system, there are customers getting on board. However, the highest value of this software is to the owners who gain transparency and speed in the leasing process.

While LeasePilot is exploring areas of opportunity for machine learning, at this point they do not anticipate this initial manual entry going away. There are certain things that people are still better for, and some other where machines take the crown, which will lead to a more efficient division of labor between people and machines.

Explorations into machine learning in the future could include using the data from all past lease negotiations in the system to predict where the negotiation will culminate, something lawyers have to do in their heads today. The tenant credit, building vacancy rates, and time of year could all be data points that lead to this type of prediction.

The system could also build in alerts for lease drafters, namely, there might be consent requirements, such as in the case of exclusive uses. These types of conflicts could be identified using simple natural language processing techniques, which in turn could be a huge benefit to both sides of a negotiation where issues will be caught earlier in the process, saving time and money. This kind of ML application is not risky and consists of a limited, well defined and very simple algorithm. Bots could also come into play to help work through these issues or to flag areas that the parties should consider. As edits are made in the leasing contract, messages could appear about alternatives or potential impacts of changing a particular clause, which could result being particularly useful as an educational tool for the tenant, as they tend not to have the same level of professional experience in lease contracts as the landlord would.

LeasePilot is also investing in blockchain to record lease terms verified by both parties, which would be more trustworthy than a report and would save time eliminating the need to read leases.

Leverton

LEVERTON

Source: Personal

Interview, Abhinav Soman, CEO

Website: leverton.ai

Year Founded: 2012

Funding: \$17.4 million (15 million euros)

Leverton is a document data extraction firm with a strong focus in the real estate world, handling lease abstracting, other real estate contracts and expanding to legal documents of all kinds.

Through the use of natural language processing and deep learning to understand sentence structure and syntax, the software achieves two important goals: firstly automating document information capture, and secondly by structuring the data in those so it is usable for analytics. This technology is useful for any field, but Leverton saw an opportunity to start out with reading contracts, specifically leases, to address the pain points there, working with investors, asset managers, property managers, law firms, and tenants. Additional real estate document types that they handle include title documents, deeds, and purchase and sale agreements. One newer area of focus is on loan documents, which they expect to be a very important addition to the software. After starting with real estate, they have expanded to other complicated document types and have continued to find pain points and work flow challenges that they believe Leverton can be helpful with. Other industries they cover include HR and Health Care, and they have gotten interesting requests to get into deeper areas such as social issues, like diversity for HR and NGOs.

In order to accomplish the contract abstracting process, the machine observes how terms are written in many contracts and learns how to differentiate and record the data in a structured form. The language comprehension element of the software uses natural language processing and neural networks, and it can manage many

different instances of language use and structuring. For example, one lease could say rent is “\$3 per SF”, where another could be written as “\$3 per annum escalating at 1%”, and yet another one could have a table with a \$3 rent listed under year 1. All of these mean the same thing, and as such the computer needs to be able to determine the rent in every instance. The accuracy is strong, but still requires human validation for quality assurance (this step can be handled by Leverton, but some customers have large teams internally to validate). In the example of a lease abstract, manually doing this task can take 6 hours or more, depending on contract length and complexity. With Leverton’s software, they are conservatively saying clients get 25% efficiency in the process. In the more recent example of processing deeds for title insurance they found that the client had 80 people who worked on extracting that data requiring takes 20-30 minutes for each one. Now that same task takes 30 seconds. These kinds of efficiencies have real implications on time, costs, and efficiency.

The platform is agnostic of document type and industry, but there are parameters they have to work within. With a new type of client or document type, Leverton requests a sample set of documents – they need to make sure there are enough examples for the machine to train on. Once they’ve agreed that it’s a type that could work with the software they then sit down with the customer to figure out how to structure the data. Following this they go through a deployment and training process which can take anywhere from 2-6 weeks. They will soon be adding a self-service version where the client can do the deployment and learning themselves, without Leverton employee involvement.

Leverton is not the only player in the document extraction space, but they have unique algorithms and systems for natural language processing and neural networks. They also built their own optical character recognition (OCR) software that uses

image recognition. The result is a more robust OCR than existing commercially available options, which have not evolved to process documents that may be in very rough shape. As the suite of documents expands, the software continues to learn new letters, words, and syntax.

When initially training the model on new contract types there is a human element, and they do not see this going away for a long time. One limitation is that in the enterprise world, you have limited sets of data, so models need to learn fast, and then rely on manual validation. Another limitation on complete automation is customers, who may not be prepared to accept a fully autonomous AI platform. There is a general concept of how a neural network works, but not total insight; this black box issue can be a problem, depending on what the customer wants to get out of it. The advancements in AI area starting slowly, but they will scale quickly. Nobody really knows whether the human element will go away entirely, perhaps more complex and long contracts will always require human legal review.

This is a global company, originally started using German technology, with capabilities across 30 languages. This also makes the software really helpful to customers who might have contracts with entities in other countries that would otherwise be hard to interpret. One interesting insight they have seen is that US leases, particularly in New York, are longer, wordier, and more complex than in Europe. This could be due to the larger amounts of “legalese” required in the US due to its being a litigious society.

The million-dollar question is, what are you going to do with the data in your system? Once the data is structured, businesses can use it to improve business processes and better understand their portfolios. However, these data analysis and visualization tools are not yet a main part of Leverton’s offerings, but will be developed over time. Some of their R&D initiatives include

developing standard analytics, for example comparing contracts to corporate policies. Additionally, system queries using AI will enable users to ask a question like, for all of my leases in Chicago how do rents compared to market?

Also important is the data security issue, which has risen to the forefront of discourse because of consumer facing products like Google, Facebook, Apple, and Amazon that got access to data before people really realized what was happening. Unlike this consumer world, enterprise has the opportunity to get it right. When operating in the US and Europe they face lots of scrutiny around access to private data. With Leverton, the customer owns the data and the platform does not aggregate the data for analytics, and they do not have that on the roadmap at this point. Ultimately, this would be a service performed by data aggregators instead. Of course, Leverton does take anonymized data and aggregate it to teach its singular “brain” with an outcome that provides a social good and progress in the ability to perform this work better over time.

For Somani, it is helpful to see how other markets outside the US work – Europeans are more interested in trying new technologies whereas Americans have less patience.

The real estate industry has been a laggard in adopting tech, but now you see the investment and seems like it is ready for change and disruption. According to Somani, they are still in the early innings of this – “our ‘brain’ is still a baby, not even a toddler yet,” but they are seeing lots of growth and more attention on this space.

Locate AI



Source: Personal

Interview, Sid Newman, Senior Vice President

Website: locate.ai

Year Founded: 2014

LocateAI is a company that provides retailers with highly detailed market information, providing insights into optimal locations, customer characteristics and behavior, competition, online sales, traffic patterns, competition, ideal co-tenants, and more. They can combine a wide range of data sets to provide insights while also taking into account a retailer's business requirements and rules, such as overlap of trade areas, revenue thresholds, etc. The goal is to help clients make real estate decisions intelligently using machine learning. It can help retailers (and potentially landlords) minimize risk, share data more efficiently, and increase sales volume. To do this they assess individual location reports and make comparisons across markets and roll data up to show info for a whole chain or set of retailers. They can achieve up to 50% more accuracy than traditional modeling in terms of projected sales and other metrics; they are very careful about how they define error

For a given block group, the software uses 160,000+ attributes and runs millions of scenarios to get an understanding of the attributes that are impactful to sales, going much further than traditional models could. It then runs millions of scenarios to get a list of attributes that are impactful on sales, going much deeper than traditional models would be able to (which usually have at most 150 attributes). The software does more than traditional machine learning by finding relationships between and within its algorithms. All of this is done at much greater speed and lower cost than traditional analytics services, and at minimal risk to the client. It only takes 2-4 weeks to get a model up and running. The model then

re-calibrates on the fly when new real time information gets added.

The data used on the platform is obtained from a variety of partners, including Esri, Weather Underground, Yelp, FBI (for crime statistics), Inrix, Factual, data.gov, CDC, OpenStreetMap, UberMedia, TR, Restaurant Trends, STI POP stats (they get data from the Post Office), and UberMedia (for massive mobile data).

They avoid the perception of machine learning and AI as a "black box" by providing transparency to the customer, doing human checks, and organizing and processing data in a way that can be tracked. Machine learning also allows for real time updates to all data and live recalibration of the models.

Locate AI has typically worked with the tenant side, but they are seeing strong interest from landlords. In order to implement the analysis for landlords they can perform similar market studies to help market space to retailers, and they can also make a comparison to other neighborhoods that the landlord might want to mimic in some way.

Motionloft

Motionloft

Source: Personal

Interview, David Wiley, Sales

Website: motionloft.com

Year Founded: 2010

The software and hardware produced by Motionloft is used to monitor pedestrian and vehicle data using sensor technology. The device they created includes a few components: cameras, internal processor, and a cellular service sim card.

Having the processor in the device allows the video image to be translated into relevant data on the fly and transmitted without needing to handle large video files or concern about privacy since all data is anonymized. The processor interprets the video based on characteristics that it recognizes base on training, for example identifying, and differentiating, pedestrian, vehicle, and bicycle traffic. The computer vision machine learning model is able to distinguishes the different modes, and can eliminate things like wheelchairs as not being bicycles

The anonymity of the data is important to clients, for example this was important in discussions with the Boston Downtown Crossing BID. The only time that video is actually recorded is during the calibration process, which requires a visual check to ensure results are producing a high-level of accuracy. The threshold is 90% accuracy before sharing the data with the customer.

Because the data is processed live, within the device, and there is no need for post-processing, so the data is instantly available to the client on a dashboard that gives clients a way to process/analyze the data, provided at an hourly level. This includes the overlay of other relevant data, such as weather. Motionloft account managers develop custom reports for each client.

CSV and JSON data are also available for the customer to be able to perform their own data modeling.

Early adopters included REITs and the commercial real estate sector as well as the public sector. Real estate investment groups want to know peak traffic hours and other data for their portfolio. They have used this type of data in the past, for example DOT traffic counts, or pneumatic tubes at entrances to shopping centers tied to an assumption about the number of people in each car. These methods are not accurate or granular. The public sector is interested in using Motionloft for retailer attraction, event tracking, and impacts of interventions and events on traffic. It's important to understand whether people are utilizing the public investments, how much traffic an event generated to enable future sponsorships, and other insights into traffic patterns.

Over the past two years or so they have been seeing stronger interest from brokers and retailers. With a shift to urban areas, retailers and city planners have to determine what makes sense in terms of urban foot traffic to support retail, which is very different than suburbs and shopping centers. With Motionloft they are able to quickly answer these questions and help retailers get comfort in entering new markets. As part of the due diligence process tenants can ask to put the sensor up in front of the window and measure capture rates (walking by vs. entering), and then assign a dollar value and compare it against a benchmark of well-performing sites.

The cost of a 30-day study is under \$2 thousand, so this can be a very good return on investment for retailers. Now that retailers are requiring this data brokers are falling in line and gathering this data as well.

Motionloft recently introduced an interior analytics product that helps retailers get insights

into what is happening inside of their stores. This includes path of travel (what direction do they go?), dwell time (how long do they stay in an area), lines (how long will people wait in line?), and heatmapping of the most heavily trafficked areas. This kind of data can be helpful for store layouts, staffing requirements, and assessment of customer service.

There is no competitor in the business who does this kind of analytics outdoors, but there are competitors that track people's movements inside stores. However, they use video based systems, which require post-processing time, and lack data granularity. Additionally, the Motionloft sensor does not require access to infrastructure because they use AT&T network. Another competitive advantage is the user experience and reporting interface.

(Wiley, 2018)

Mynd



Source: Personal

Interview, Colin Wiel, Chairman, CTO, and Co-Founder (Wiel, 2018)

Website: mynd.co

Year Founded: 2016

Funding: \$35.6 million

Mynd is a residential property management company that uses technology for an enhanced management platform and tenant experience. Property management is an underserved industry with limited technology and lots of friction, but many decisions are data-driven. Ultimately many functions can be done better by an algorithm than a human. Mynd currently manages properties in California, and will soon be expanding to other markets.

Co-founders Colin Wiel and Doug Brian previously started Waypoint Homes, a single family residential rental company, something people had never considered to be a scalable enterprise. The company eventually went public with Starwood and later merged with Blackstone's Invitation Homes. When Colin and Doug decided to transition to Mynd, they recognized that most emerging companies now are "asset light," with limited ownership of physical spaces or products and started looking at the property management space.

Colin thinks of residential property management in terms of "small" assets consisting of single family and up to 50 unit buildings, and "large" assets with more than 50 units. Small assets are hard to manage because there can't be a single dedicated staff person. There is a big opportunity here; the annual revenue for all of these units is \$433 billion annually, but the market is highly fragmented and nobody is operating at a large scale.

The first step for the company is to build out the basics to do property management well and at a larger scale than typically seen. One technology they've adopted early on is smartlocks, which allows for "self-showings," access for maintenance, and other efficiencies.

Over the next few years the goal will be to systematize and automate the things that humans typically need to do using machine learning and AI, and elevate people in the company to higher levels of decision making. Some potential uses for machine learning include:

- vendor selection
- optimizing rent levels
- prediction of when maintenance will be required
- Integration with IoT

These areas could work using support vector machines, linear regression, and other proprietary algorithms. As the robust dataset from properties under management grows they will have the ability to perform powerful analytics.

Nano Global and
Nano Vision



Source: Personal

Interview, Aaron Papermaster, Product and Business Development (Papermaster, 2018)

Website: nanoglobal.com and nanovision.com

Year Founded: 2013

Nano Global aims to address major world problems through nano scale sensing and analytics of the most granular data available. Specifically, they are working on what is happening inside the walls of buildings with a digital infrastructure for research in the health field. As noted on the website: "our technology makes real-time molecular data addressable and scalable to ultimately redefine how human and environmental health challenges are conquered" ("Nano," 2018). They are deploying highly accurate sensors in health care facilities to start, and then are expanding to hospitality and other uses. While they are initially providing hardware, smaller and more accurate devices than others on the market, their long-term intention is to be focused on the data side.

The idea comes from the public health risks posed by poor air quality. On average, approximately 90% of your time is spent indoors, whether at home, work, or school. Chronic exposure to diseases or toxins in the air in any of these spaces can lead to significant degradation of health. One example was a study linking a certain level of particulate matter in the indoor environment to developing diabetes. As we are beginning to understand with the study of epigenetics, even our kids can be impacted by exposure from a previous generation.

Today there are plenty of devices that can give you interesting insights, but the information as provided is not really actionable or that impactful. For example, you may learn that the CO₂ content of the air is high, so you should get a plant, but

can't take much action beyond that. However, if you can start linking chronic exposure to toxins to particular diseases we can make really actionable insights that impact health in profound ways. In the past, studies that were done in isolation and siloed have been unable to produce these kinds of insights. Nano envisions a world where you can capture this information pervasively throughout the world.

The sensor technologies being developed also allow for detection of air borne diseases. Today, if someone walks into a hospital waiting room or a hotel lobby you can't tell if they might have some kind of disease. With new sensors it could get to the point you would know as soon as they walk in the door. This would be possible because there are trace amounts of molecules in the air when someone has an illness. This kind of analysis resembles the problem of finding a needle in a haystack, but with machine learning this task could be possible. With AI comes the ability to find anomalies by comparing the sensor readings to healthy air while filtering out lots of noise.

These computations are done in the device, not in the cloud. With AI and ML even the timing and frequency of measurements can automatically be adjusted within the device. We can also identify correlations the environment at work, and home, and layer in genetic data, lifestyle, and health history. This is an example of the convergence of several technologies: AI, cutting edge sensor technology, and edge computing technology.

Navigator CRE

Source: Personal

Interview, Taylor

Odegard, CIO, Founder, and Chairman

Website: navigatorcre.com

Year Founded: 2015



Navigator CRE founder Taylor Odegard studied real estate finance and development and realized along the way that investors are using antiquated technologies in a trillion dollar industry, and methods were still very "provincial." In 2008-09 he started gathering and testing his ideas for a cloud based platform that would combine maps, spreadsheets, and dashboards with granular data on building types, features, etc. He then developed early versions of Navigator during his time at CBRE, leading their west coast capital markets team. Clients there would ask for his model by name. The interactive nature of the platform allowed clients to ask questions and get granular answers in the moment.

The Navigator CRE service is intended to provide analytics for all parts of the real estate field. The platform is data agnostic, they are not a data provider. Navigator integrates with other platforms including Salesforce, Yardi, Argus, MRI, and many others. It acts as a repository for all kinds of data where owners can upload their data and then the software maps everything and automatically generates analytics, all secured, private, and protected. Asset manager can take it everywhere and have all the stats they need on a portfolio. Investors can virtually tour markets. Teams can collaborate on the cloud based smart-grid engine. The technology has achieved acclaim and recognition for providing actionable intelligence.

The dashboards are fully customizable for each client, and a team of business intelligence professionals builds out dashboards for each

client. The team does a lot of data processing, data ingest, etc. Often the dashboard will show anomalies that portfolio operators might not have seen in the past without this resource. As the client base grows and they get more insights into customer requirements, this will continue to evolve.

They are also looking into additional integrations with platforms like Microsoft's AI. Their AI functionality can be embedded in the system so that the AI can do predictive analytics without the Navigator team needing to get involved; an active algorithm constantly evolves and pulls in the data that is needed. In addition, ML could be used to write a narrative about what is happening in a portfolio. They are also thinking about active forecasting by working with other companies that have valuation data. They are also interested in integrating with IoT as building learn to operate through automation and integrate with a "hive mind" system. However, Odegard cautions that it is important to limit this to actionable insights.

Opendoor

Website:

[opendoor.com](https://www.opendoor.com)

Opendoor

Year Founded: 2014

Funding: \$645 million

Opendoor allows homeowners to sell their properties online quickly and easily, purchasing them directly, and then finds buyers through their online platform. They run all-day open houses where potential buyers can use an app to gain access to the property after they provide identifying information. They are currently operating in Atlanta, Charlotte, Dallas-Fort Worth, Las Vegas, Minneapolis-St.Paul, Nashville, Orlando, Phoenix, Raleigh-Durham, San Antonio, and Tampa. ("Opendoor," n.d.)

The service helps homeowners to quickly exit their property with a transparent fee structure, based on risk, to cover Opendoor's cost of reselling the asset. The seller benefits because they do not have to pay \$300 or more to get an appraisal, they have a guaranteed offer, and no delay in the transaction. Since data is public for these sales the accuracy of pricing can be verified and over time the fair pricing of the model will be apparent. There are also many pain points addressed on the buyer side. This includes a trial period with a money back guarantee and warranty on equipment and other items that have been updated during the sales process. (Wong, 2017)

The business model requires various forms of data science, broadly categorized by a data scientist there as "ingesting and organizing data," home value price prediction and market risk, quantification and mitigation of risk, and collecting all of the data gathered in a "data warehouse." He states that he has "found data science at Opendoor interesting because it's not the "grab as much data as you possibly can, then process it at huge scale" problem I'm used to

hearing about.” Rather, the platform gathers important data points and an automated system “grabs the required features, trains the model, and evaluates how well the model did against performance metrics,” allowing them to iterate quickly. (Zirbel, n.d.)

In order to offer this service they rely on automated valuation modeling to accurately price homes. Opendoor uses machine learning in myriad ways to achieve these valuations. They use deep learning to identify important value features based on property photos. In addition, they find “wide” data points on properties including features, such as whether it has a pool, is on a cul de sac, has a nice front yard (and curb appeal), neighboring properties, and more. However, that data is sparse, or in other words incomplete for each property. Market data is used to compare asset sales across time as well. They are also building recommendation engines into their selling side, helping purchasers find the right home for them. (Wong, 2017)

Placeful

Source: Personal Interview, Daniel Fink and James Scott, Co-Founders

Website: placeful.io

Placeful

Year Founded: 2017

Started by a group of MIT alumni and researchers, Placeful aims to address the deficit of 6 million housing units in the United States through a technology-enabled methodology to identify sites and quickly optimize a design for residential development. They will achieve this using AI to optimize the space usage and to assess the value and other important factors of site acquisition. They will attempt to answer the question: where should we build and what should we build?

The idea is to match sites with pre-designed building components to a particular site to enable housing construction at a lower cost. In a typical residential building some components are fixed, such as staircases, which have to be a certain dimension. Whereas other components are not pre-defined, like room size. Working with a pre-defined kit of components the software will be able to perform “generative design,” to create the optimal building for each site, a remarkable form of AI being applied to physical form. The technology will use components of machine learning, which the team considers to be, at its core, all about optimization. This particular case is an example of a *bin packing problem*, or a problem where the optimal configuration for a set of volumes needs to be determined.

The design will also iterate and offer options with different trade-offs that will inherently alter the end physical form, for example: should you build out all of the available square feet allowed under zoning if it costs more? Or, should you break a zoning regulation and go to the expense and risk of requesting a variance? These types of second-

order impacts can be assessed quickly and easily once the impacts are made clear by the model.

To start off this project, however, the team will have to generate their own designs to teach the model in order to get around the challenge that there is very little data on existing development and design is very idiosyncratic. The firm is forming a team of architects, engineers, and contractors, and will work with limited partners to develop their projects. They intend to target vacant lots and tear-downs so they can bring housing to underutilized land. The form will be “triple deckers,” a common residential typology that can be executed quickly.

The cost of construction can be estimated relatively easily given the panelized system of components, and most of the other required data is readily available. However, there is no data source for land sales in the cities. The land market is a crucial component to understand, but it is a very opaque market.

With some experimentation and a methodology for assessing land values the team will have the tools to create a system that quickly automates site analysis and schematic design, saving time and expense.

(Fink, 2018; Scott, 2018)

Real Capital
Analytics
(RCA)

REAL CAPITAL
ANALYTICS®
Transactions. Trends. Tools.

Source: Personal Interviews, Willem Vlaming, Analyst, and Ofer Goldstein, former Data Scientist

Website: rcanalytics.com

Year Founded: 2000

The RCA platform leverages partnerships with industry titans and premier research institutions (like MIT) to provide in-depth market level insights. They typically have created indices and other tools based on cutting edge econometric models. I spoke to Willem Vlaming, an analyst at Real Capital Analytics, and Ofer Goldstein, a former employee there. In these conversation I learned that the firm has explored various applications of machine learning, and are now in the early stages of a valuation project. This is not typical for RCA, which is typically a data and reporting platform, not doing projections.
(Goldstein, 2018; Vlaming, 2018)

The new valuation project is aimed at estimating transaction pricing for all properties, even those missing data points. They have been able to do this preliminarily exclusively using RCA data, including transactions, lender activity, and RCA generated price indexes and quality scores. RCA does not use leasing data in their analysis. Some of the most important value indicators identified so far are location (they integrate with mapping), space/size, and age and renovation year. In order to enhance the dataset, which pulls data from multiple sources they do what they call *dataset matching* using techniques such as *string matching* and *record linkage*. (Vlaming, 2018)

Goldstein also noted that in order to collect additional data they employ machine learning techniques to scrape SEC data for CMBS transactions and link the information gathered to their database. (Goldstein, 2018)

In one part of their analysis, they employ the *K Nearest Neighbor* algorithm for clustering to find properties that are similar based on location, size, age, etc., and these qualities are weighted depending on the factors the particular client finds important. This is also used for RCA's website to find comparables for clients.

Some of the challenges identified included overfitting and the fact that the available algorithms tend to be very general, and not built specifically for RE due to the sources of those algorithms. But, of course the data is really the key and the hardest work.

Reonomy



Source: Personal

Interview, Ismail

Pathan, Customer Success and David Vigilante,
Product Manager

Website: <https://www.reonomy.com/>

Year Founded: 2013

Reonomy provides 2 real estate products: a consumer end web application for property and owner data and an enterprise data service. The web application gives users access to information about the ownership of companies and properties, getting behind the SPE LLCs and determining who the key stakeholders of the company might be. This is combined with building data on sales transactions and building characteristics. The enterprise product provides similar information but integrates with customer data using APIs.

Reonomy uses machine learning in data gathering and record linkage. They need to match company names with the proper LLCs or reporting entity, and link it other data sources, such as Secretary of State data. They have achieved mid-90's accuracy. Similarly, they use these techniques to combine different data sources with varying address names for same property observations. Machine learning is also used for sales comps to pull 5-10 properties that are relatively similar to a subject property with recent trades.

The product was initially targeted towards commercial lenders, brokers, and others in NYC. However, when they rolled out the national product in 2017 they found that there was a large group of service providers that needed access to property owners. For example, a solar provider might identify opportunities for commercial roof installations and will need to know who to reach out to discuss the opportunity. In other words, vendors are using it as lead generation and prospecting tool. Now they are working to figure

out how to offer better service different types of users. The firm has been inundated with various interests who want access to commercial real estate owners, even a drone operator.

Data sources include Secretary of State data, assessor records, and other public records. The availability of this information varies by county. There are almost 3,000 counties in the US, and the availability of data comes down to how tech savvy the county is. Some regions still do not even have their records digitized. They are also working with some private vendors for contact details and tenant information.

Reonomy has also found that a lot of enterprise clients are trying to build their own infrastructure to utilize the large amounts of data in their systems. However, it tends to be spread across databases. In these cases they can use the Reonomy data resolution API to help consolidate and clean up their records. Then they can use property search API to pull in additional data points, bringing it all together. Brokers and financial institutions have so much data they don't know how to use (and there is no need for them to share data, they can leverage what they have internally by combining with outside sources).

While continuously working to improve the AI they have, Reonomy is looking into other new areas for ML use. There could be ways to query the system for searches like the locations for all stores of a certain type, or the number of properties owned by a particular person. However, the biggest challenge is finding enough engineers and combining their knowledge with commercial real estate, as they do not have this background.

REview Analytics

Source: Personal Interview,
Michael Pearce, Founder



Year Founded: 2018

REview is in the early stages of business planning, coming out of the MIT DesignX incubator program. Founder Michael Pearce came from a background in real estate at Divco where he saw the difficulties of competing for bids due to the opacity of how others in the market were valuing assets. Without a basic suite of data and analytics tools real estate professionals have to struggle to keep up with market trends, from what tenants are looking for, to capital market trends, to asset management needs. REview is working on refining its plans for providing a platform that will replace basic market research at micro and macro levels, including hiring data, start-up data (location, focus, hiring), and information from public databases such as BLS, the Census Bureau, US PTO, and the IRS (for data on migration). They are also contemplating the possibilities for how that can scale. With a powerful data tool, the user interaction with that database can become a source of data and insights in itself when machine learning and artificial intelligence are brought in.

The idea for REview software started with anonymizing data from user input data such as Argus models to assess markets using very detailed data on each asset. However, there is a lot of skepticism from the industry about providing this kind of data, which is regarded as highly proprietary.

(Pearce, 2018)

Ten-x/Auction.com

Source: Personal
Interview, Yiqun Wang



Website: ten-x.com and
auction.com



Year Founded: 2007

Ten-X Commercial and Auction.com are online marketplaces for buying and selling real estate assets. As a marketplace, Ten-X Commercial co-lists assets with brokers, consolidates financials and due diligence, orders third party reports (like title and phase 1 environmental), and more for each asset. By bringing these services into an online platform, it broadens the market for commercial property buyers.

In parallel, Auction.com provides services to sellers of distressed residential assets, such as banks and institutional investors. The platform can handle everything from the marketing of a distressed asset, to the sale of an entire distressed portfolio. They apply data science to assess demand and guide pricing, which requires a unique methodology given that they are dealing with distressed assets and often no interior access to the property. (Wang, 2018)

Truss



Source: Personal Interview, Thomas L. Smith, Co-Founder

Website: truss.co

Year Founded: 2016

Funding: \$9 million

Truss is a commercial real estate online leasing marketplace with built in analytics. The platform addresses the problems faced by small and medium businesses looking to lease space: there is a lack of transparency and there is friction from intermediaries in the business transactions, who are not interested in these smaller deals. Of the four co-founders, only one came from a commercial real estate background, the other are from tech and FinTech industries and they take an approach that breaks many of the rules of traditional real estate deals.

They started by targeting the office space market for small and mid-sized companies. Co-founder Thomas Smith explained that of the many lease transactions in every major market only a fraction of them are large tenant, but they get all the attention. For example, there are roughly 14,000 office lease transactions in Chicago annually, but only 500 of those leases drive 2/3 of the tenant representation fees in that market. This situation leaves many transactions underrepresented or unrepresented. Once this platform was developed for office deals, it was relatively easy expand to retail, and they recently announced their entry into the industrial space.

For small and medium sized tenants operating with no access to a public MLS and no real estate specialist on staff, an inexperienced person has to lead the space search, but doesn't get much help or attention from the broker. The standard process for leasing office space looks like this: the broker will do a phone interview, then run a

search on a platform that does not include pricing information for most properties and tends to lack other information and photos, and then they go on a tour. To fix these problems, educate clients, and save time and effort, Truss offers a service where they have licensed brokers in every market they serve to conduct property tours, but they generally operate using 95% technology, and only 5% human interaction. The goal is to put the tenant in control of the leasing experience, allowing them to browse availability in the market, know if they can afford it, take a virtual tour, and do it all at whatever time is convenient.

The artificial intelligence component is a bot named Vera that does more than chat. A tenant registers on the site, for free, and Vera gathers information on the space requirements, including the number of people and growth projections, pin-points on a map of preferred locations, the tenant's industry, space layout preferences, and other qualities they might prefer (for example would you rather pay for more space or nicer amenities?). The service then finds matching properties in the available listings and the client can then pick out a short-list, set up tours, and go through the lease negotiation process all on the platform. Vera can assist in other ways, for example the platform can guide tenants to what is a marketable rent ask, and provide coaching so they do not miss out on a deal by going too low. The system learns what terms are acceptable in the market based on executed leases and anecdotal evidence from sources like Compstak (and the platform knows how these anecdotes trend to actuals so can adjust for that).

In order to list a property, Truss requires full transparency into lease costs upfront (net vs gross, etc.), and then they translate that into a total monthly rent and expense number that makes sense to the tenant as a business. The space listing also provides information about the building from various sources, such as transit and walk scores. Additionally, Truss has partnered

with Matterport to provide 3D virtual tours of each property as a way to give the tenant detailed insight into a property before wasting time and energy on tours.

As the tenant rep, Truss provides full transparency into leasing commissions they are getting from each deal, even though there are no disclosure rules for CRE brokerage. Plus, they give the tenant 30% of the leasing commissions they receive as tenant rep.

Truss has been able to achieve listings for 95% of the availability for small to medium sized spaces in each market they serve because brokers and owners get access to a marketing channel for non-accretive business for free. Co-working spaces are also interested in using the platform as it gives them added exposure. Brokers are further endeared to them because they refer out big deals to the existing tenant rep broker community. For Truss, there is more value in smaller deals, which bring in more data points.

The data gathered on the platform can lead to many insights beyond reasonable rental rates for tenants. While the platform provides a great service to the tenants in their space search, it also allows Truss to gather important market insights in the process. By tracking clients' preferences, short-list selections, lease negotiations and ultimate lease terms, they have access to very detailed market insights based on actual demand, like an exact pinpoint on the map, rather than where tenants end up settling. The data can be anonymized and sold back to owners and investors to give them actionable insights into markets and trends, and analysis can be packaged in reports and shared in workshops with landlords and brokers.

One example of an insight is optimized pricing strategies for landlords, helping them to understand whether a lower price would have been better to lease up the space faster. By watching lease term negotiations they've

discovered that where they start doesn't matter too much, they end up in the same place anyway. Therefore, landlords are better off marketing at a lower price because they'll end up at that lower price in the end.

One area of machine learning that they are utilizing is geo-spatial data analytics techniques that they have adapted from mining industries. In mining, a line of gold might be an indication of what you might find along the terrain. This can be translated to a transit line, which provides clues about what might be found nearby. Using these techniques they can understand demand based on neighbors and other local qualities.

From the locational data gathered as tenants pinpoint their preferences and execute leases they can create a heat map of demand. Property investors can use this insight to track demand movements like a weather pattern, and Truss has generally observed a traceable path of shifting demand. In the future they will layer in a predictive component.

Future plans for the platform include tenant experience and feedback on the process and occupancy experience, making them the Yelp of CRE. Traditionally, property management has been reluctant to ask tenants how things are going to avoid complaints, and owners are blind to issues at the property.

Additionally, they will be rolling out new feature where space definition from Matterport can be used to generate design options to bridge design questions that tenants may have. Space "skinning" can be done using technology from a company called BlockVue, and furniture from vendors like Steelcase and Knoll can be dropped in to "virtually stage" the space. A lot of this will also be fully or partially automated in the future to enable this at a higher volume scale, BlockVue has already automated a lot of the process.

In conclusion, Truss provides tenants with the ability to get true information about market dynamics, resolves the frictions created by misaligned interests from brokers, provides granular insights into market demand and lease deals, and has the opportunity to transform the leasing business.

(Smith, 2018)

Zillow



Website: [zillow.com](https://www.zillow.com)

Year Founded: 2005

Stock Symbol: NASDAQ:Z

With their Zestimate, Zillow has made headlines and ruffled feathers. Nobody wants to be told that an algorithm downgraded their home's value by 10% last month. This type of complaint and others have led to several law suits against the company, which were dismissed by a federal judge in May of 2018. A part of the suit "alleged that Zillow systematically engages in a confusing, unfair and deceptive marketing scheme that impairs homeowners and sellers in the sale of their houses" (Harney, 2018). However, the Judge determined that the site was not claiming to provide anything more than an estimate. A description of the Zestimate available on the Zillow website states, "the Zestimate home valuation is Zillow's estimated market value for a home, computed using a proprietary formula. It is a starting point in determining a home's value and is not an official appraisal. The Zestimate is calculated from public and user-submitted data" ("Zillow," n.d.).

Zillow gets some of its machine learning for estimating home values from open competitions. Last year, they hosted a competition called Zillow Prize with a \$1.2 million prize on Kaggle. As described on the competition site,

““Zestimates” are estimated home values based on 7.5 million statistical and machine learning models that analyze hundreds of data points on each property. And, by continually improving the median margin of error (from 14% at the onset to 5% today), Zillow has since become established as one of the largest, most trusted marketplaces for real estate information in the U.S. and a leading

example of impactful machine learning” (“Zillow Prize: Zillow’s Home Value Prediction (Zestimate) | Kaggle,” n.d.).

Their automated valuation model is one of the most well-known in the market and is often used

as a point of comparison for other types of valuation services. According to real estate data provider, Foxy AI, founder, Zillow is the biggest provider of AVMs, and they use regression tree analysis techniques, likely in combination with other techniques (Vomero, 2018).

Glossary

The following definitions are the author's own interpretation of each of these concepts as used throughout the paper as they pertain to machine learning.

target variable/target vector: the variable that a model is trained to predict; these can be categorical variables if used for classification or continuous variables if a regression problem

training set: an initial dataset including input variables (and target vectors in the case of supervised learning) used to train the algorithm

holdout set: commonly called "testing" data, it represents a subset of the data used to estimate the model's performance post training

bin packing problem: a problem used in computational complexity theory referring to the problem of whether objects of different volumes, shapes, sizes, weights, etc. can fit into a pre-determined number of containers, or bins (considered to be a combinatorial NP-Hard problem)

dataset matching/record linkage: identifying and aggregating of data archives that match entries under multiple names or identifiers to the same real-world entity

string matching: characterizes the process of matching 2 different but similar pieces of data (sometimes even entire databases) by having the computer group them together using text elements

feature engineering: the procedure of utilizing domain knowledge regarding the data to generate new dataset elements to enable machine learning algorithms

boosting: a machine learning collective meta-algorithm mainly for reducing bias, and also variance in supervised learning.

sequential bootstrapping: consists of a sequential training method for developing bootstrap aggregated neural network models, therefore trained sequentially

vector quantization and learning vector quantization: compression techniques that reclassify data points based on proximity to other points; this uses the results of clustering algorithms to simplify the data in each cluster to all be the same; it is an example of lossy compression, where information is lost in the process (Bishop, 2006, p. 429)

Works Cited

- 8 Fast Growing Technologies. (2018). Retrieved from <https://www.ificlaims.com/rankings-8-fast-growing.htm>
- 9 Early-Stage Enterprise AI Startups To Watch. (2018). Retrieved June 17, 2018, from <https://www.cbinsights.com/research/early-stage-enterprise-artificial-intelligence-startups/>
- Affectiva. (n.d.). Retrieved June 22, 2018, from <https://www.affectiva.com/#undefined>
- AirWorks. (n.d.). Retrieved June 25, 2018, from <http://airworks.io/#home>
- Alexa for Business. (n.d.). Retrieved June 18, 2018, from <https://aws.amazon.com/alexaforbusiness/>
- Beckerman, M. (CRE Tech). (2018, June 7). Personal Interview.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. (B. Jordan, Michael; Kleinberg, Jon; Scholkopf, Ed.), *Information Science and Statistics*. Springer. <https://doi.org/10.11117/1.2819119>
- Brinser, C. (Turner Construction). (2018, May 14). Personal Interview.
- CB Insights. (2018). Top AI Trends To Watch In 2018, 30. Retrieved from https://www.cbinsights.com/reports/CB-Insights_State-of-Artificial-Intelligence-2018.pdf
- Cearley, D. W., Burke, B., Searle, S., & Walker, M. J. (2017). Top 10 Strategic Technology Trends for 2018. Gartner. Retrieved from https://www.gartner.com/doc/reprints?id=1-4MKEN3N&ct=171220&st=sb?utm_campaign=WREC_180102_Q1_AC1121_Top%2010%20Technology%20Trends%20for%202018.02.Converted&utm_source=marketo&utm_medium=email&utm_content=main-cta-button&refcode&mkt_tok=eyJpljoiWm1Rd1lqU
- Chegut, A. (MIT). (2018, June). Personal Interview.
- Chegut, A., Eichholtz, P., & Kok, N. (2017). *The Price of Innovation: An Analysis of the Marginal Cost of Green Buildings (working paper)*. MIT : Center for real estate.
- Choosing the right estimator. (n.d.). Retrieved June 16, 2018, from http://scikit-learn.org/stable/tutorial/machine_learning_map/
- Columbus, L. (2018). Roundup Of Machine Learning Forecasts And Market Estimates, 2018. *Forbes*. Retrieved from <https://www.forbes.com/sites/louiscolumbus/2018/02/18/roundup-of-machine-learning-forecasts-and-market-estimates-2018/#3b0f9d842225>
- Construction Costs Soar in May as New Tariffs Threaten to Worsen Cost Squeeze, Lead to Delays. (2018, June 15). *Boston Real Estate Times*. Retrieved from <http://bostonrealestatetimes.com/construction-costs-soar-in-may-as-new-tariffs-threaten-to-worsen-cost-squeeze-lead-to-delays/>
- Conway, A. (DataRobot). (2018). Personal Interview.
- Dalgaard, L. (2018, January 23). Doxel. Retrieved July 2, 2018, from <https://a16z.com/2018/01/23/doxel/>
- Data Robot AI Experience. (2018). Boston.
- Definition - What is geospatial analysis? (n.d.). Retrieved June 17, 2018, from <https://whatis.techtarget.com/definition/geospatial-analysis>
- Dhandapani, C. (CBRE). (2018, July). Personal Interview.
- Directly connecting borrowers and lenders Looking for value in data. (2018). *DataRobot*. Retrieved from https://www.datarobot.com/wp-content/uploads/2018/02/Case_Study_Zidisha_final_web.pdf
- Doxel AI - Artificial Intelligence for Construction Productivity. (n.d.). Retrieved June 18, 2018, from <https://www.doxel.ai/>
- Dvele. (n.d.). Retrieved June 18, 2018, from <https://www.dvele.com/>

- Dvorsky, G. (2018, July 12). This is the Most Aerodynamic Bike, According to AI. *Gizmodo*. Retrieved from <https://gizmodo.com/this-is-the-most-aerodynamic-bike-according-to-ai-1827546083>
- ENR. (n.d.). ENR Construction Economics. Retrieved June 18, 2018, from <https://www.enr.com/economics>
- Everitt, T., & Hutter, M. (2018). Universal Artificial Intelligence. In H. A. Abbass & D. J. Reid (Eds.), *Foundations of Trusted Autonomy*. Springer.
- Finegold, E. (2018, July). Personal Interview.
- Fink, D. (Placeful). (2018, June 6). Personal Interview.
- Francke, M. (Ortec Finance). (2018, June 7). Personal Interview.
- Fromm, J. (2017, July 12). How Startup Lemonade Is Redefining Insurance For Millennials. *Forbes*. Retrieved from <https://www.forbes.com/sites/jefffromm/2017/07/12/how-startup-lemonade-is-redefining-insurance-for-millennials/#6df0d8c611b0>
- Ganes, K. B. (2018). What frustrates Data Scientists in Machine Learning projects? Retrieved June 9, 2018, from <https://towardsdatascience.com/what-frustrates-data-scientists-in-machine-learning-projects-3398919a7c79>
- Gaw, D. (2018, June 1). The Future of Construction Site Automation. *ENR*. Retrieved from <https://www.enr.com/articles/42837-the-future-of-construction-site-automation>
- Geltner, D. (MIT). (2018). Personal Interview.
- Goldberg, Y. (2015). A Primer on Neural Network Models for Natural Language Processing. *ArXiv.Org, cs.CL*, 726. <https://doi.org/10.1613/jair.4992>
- Goldstein, O. (DataRobot). (2018, April 19). Personal Interview.
- Graham, E. (Crowd C. (2018, June 26). Personal Interview.
- Grimes, S. (2005, July 2). Structure, Models and Meaning. *InformationWeek*. Retrieved from <https://www.informationweek.com/software/information-management/structure-models-and-meaning/d/d-id/1030187>
- Harney, K. (2018, May 18). What to make of the dismissed class-action lawsuit over Zestimates. *The Real Deal*. Retrieved from <https://therealdeal.com/2018/05/18/lawsuit-dismissed-against-zills-zestimates/>
- Henn, A., Römer, C., Gröger, G., & Plümer, L. (2012). Automatic classification of building types in 3D city models. *GeoInformatica*, 16(2), 281–306. <https://doi.org/10.1007/s10707-011-0131-x>
- Hong, L. (DealPath). (2018, June 14). Personal Interview.
- Ibanez, L. (2018, June 15). How AI for AI Will Change the Industry. *Freddie Mac*. Retrieved from http://www.freddiemac.com/perspectives/loretta_ibanez/20180615_artificial_intelligence.html
- Introducing The Topos Similarity Index and [x] Everywhere. (2018, November 16). *Medium*. Retrieved from <https://medium.com/topos-ai/introducing-the-topos-similarity-index-and-x-everywhere-82fce1fb367>
- Isaacs, N. (Bowery Residential). (2018, June 7). Personal Interview.
- Isazi Consulting. (n.d.). Retrieved July 1, 2018, from <http://www.isaziconsulting.co.za/machinelearning.html>
- Jolliffe, I. T. (2002). Principal Component Analysis, Second Edition. *Springer*. Retrieved from [http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe_I._Principal_Component_Analysis_\(2ed.,_Springer,_2002\)\(518s\)_MVsa_.pdf](http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe_I._Principal_Component_Analysis_(2ed.,_Springer,_2002)(518s)_MVsa_.pdf)
- Jones, B. (2017, November 19). IBM and MIT are collaborating to ensure that we develop A.I. technology safely. *Digital Trends*. Retrieved from <https://www.digitaltrends.com/computing/ibm-and-mit-ai-partnership/>
- Khan, P. (CoStar), & Lupton, S. (CoStar). (2018, March 15). Personal Interview.

- Kok, N., Koponen, E.-L., & Martinez-Barbosa, C. A. (2017). Big Data in Real Estate? *The Journal of Portfolio Management*, (Special Real Estate Issue | 8th Edition).
- Leach, J. (Leverton). (2018, June 6). Personal Interview.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436. Retrieved from <http://dx.doi.org/10.1038/nature14539>
- Leinberger, C. B., & Loh, T. H. (2018). Catalytic development: (Re)creating walkable urban places. Retrieved from https://www.brookings.edu/wp-content/uploads/2018/05/brookings-180420-catalytic-development-paper_may-2018-final.pdf
- Lemonade. (n.d.). Retrieved June 18, 2018, from <https://www.lemonade.com/>
- Lerner, Y. (JLL Spark). (2018, June 12). Personal Interview.
- Leviathan, Y. (Google), & Matias, Y. (Google). (2018). Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. Retrieved June 23, 2018, from <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>
- Levy, N. (2018, October 2). Realtors vs. robots: Redfin CTO says artificial intelligence shouldn't replace real estate agents. *GeekWire*. Retrieved from <https://www.geekwire.com/2017/realtors-vs-robots-redfin-cto-says-artificial-intelligence-shouldnt-replace-real-estate-agents/>
- Lewis, G. (Honest Buildings). (2018, May 18). Personal Interview.
- Li, H. (2017). Which machine learning algorithm should I use? Retrieved July 1, 2018, from <https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/>
- LocateAI. (n.d.). Retrieved June 24, 2018, from <https://www.locate.ai/>
- Loizos, C. (2018, March 8). This startup just raised \$5 million to automate the clunky real estate appraisal process. *Tech Crunch*. Retrieved from <https://techcrunch.com/2018/03/08/this-startup-just-raised-5-million-to-automate-the-clunky-real-estate-appraisal-process/>
- Ma, Y., Zhang, Z., Ihler, A., & Pan, B. (2018). Estimating Warehouse Rental Price using Machine Learning Techniques, 13(April), 235–250.
- Machine Learning | DataRobot Artificial Intelligence Wiki. (n.d.). Retrieved June 17, 2018, from <https://www.datarobot.com/wiki/machine-learning/>
- Machine Learning 101 The Kernel Trick. (2018). Retrieved June 16, 2018, from https://www.youtube.com/watch?v=7_T9AdeWY3k
- McGill, C. (Enertiv). (2018, June 12). Personal Interview.
- Moyer, C. (2016). How Google's AlphaGo Beat Lee Sedol, a Go World Champion. *The Atlantic*. Retrieved from <https://www.theatlantic.com/technology/archive/2016/03/the-invisible-opponent/475611/>
- Nano. (2018).
- Newman, S. (LocateAI). (2018, June 5). Personal Interview.
- Odegard, T. (Navigator CRE). (2018, June 12). Personal Interview.
- Opendoor. (n.d.). Retrieved June 30, 2018, from <https://www.opendoor.com/>
- OpenSpace. (n.d.). Retrieved June 18, 2018, from <https://openspace.ai/>
- Panetta, K. (2017, August). Top Trends in the Gartner Hype Cycle for Emerging Technologies, 2017. *Smarter With Gartner*. Retrieved from <https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/>
- Papermaster, A. (Nano). (2018, July). Personal Interview.

- Parr, T., & Howard, J. (n.d.). How to explain gradient boosting. Retrieved June 23, 2018, from <http://explained.ai/gradient-boosting/index.html>
- Pearce, M. (REview). (2018, June 12). Personal Interview.
- Pentland, A. (2015). Data-Rich Cities. In *Social Physics*. Penguin Books.
- Popken, B. (2018, May 10). Google sells the future, powered by your personal data. *NBC News*. Retrieved from <https://www.nbcnews.com/tech/tech-news/google-sells-future-powered-your-personal-data-n870501>
- Prado, M. L. De. (2018). THE 10 REASONS MOST MACHINE LEARNING FUNDS FAIL. Retrieved from www.QuantResearch.org
- Product | OnSiteIQ. (n.d.). Retrieved June 22, 2018, from <https://onsiteiq.io/product/>
- PROP. (n.d.). Retrieved June 18, 2018, from <http://www.prop.uk.com/about/>
- PSKF. Interview: How AI And Data Analysis Bolster The Real Estate Business, pskf (2018). Retrieved from <https://www.psfk.com/2018/04/bobby-goodman-truss-ai-data-analysis-real-estate.html>
- Rutzen, M. (Enodo). (2018, June 6). Personal Interview.
- Sadler, M. P. (2018). *Data Science and Advanced Analytics : An integrated framework for creating value from data*.
- Salesforce Einstein. (n.d.). Retrieved June 18, 2018, from <https://www.salesforce.com/products/einstein/overview/>
- Sargent, J. (2018). Databricks releases MLflow, runtime for ML and Databricks Delta at Spark + AI Summit. Retrieved June 10, 2018, from <https://sdtimes.com/ai/databricks-releases-mlflow-runtime-for-ml-and-databricks-delta-at-spark-ai-summit/>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Scott, J. (MIT). (2018, April 10). Personal Interview.
- ShopoGraphics. (n.d.). Retrieved June 22, 2018, from <https://www.forumanalytics.com/solutions/shopographics/>
- Shu, C. (2018). Skyline AI raises \$ 3M from Sequoia Capital to help real estate investors make better decisions.
- Simonson, K. (Associated C. C. of A. (2017). Construction Costs Outlook. Retrieved June 18, 2018, from <https://www.naiop.org/en/Magazine/2017/Winter-2017-2018/Business-Trends/Construction-Costs-Outlook>
- Skyline. (n.d.). Retrieved June 24, 2018, from <https://www.skyline.ai/platform.html>
- Smith, T. (Truss). (2018, June 8). Personal Interview.
- Snider, D., & Harris, M. (2018, February 13). The Future Of Real Estate Tech: How We Got Here And What's Next In An Exploding New Ecosystem. *Forbes*. Retrieved from <https://www.forbes.com/sites/valleyvoices/2018/02/13/future-of-real-estate-tech/#79d3460d1c13>
- Somani, A. (Leverton). (2018, June 14). Personal Interview.
- Spitzen, I. (LeasePilot). (2018, July 10). Personal Interview.
- States Title | Crunchbase. (n.d.). Retrieved June 18, 2018, from <https://www.crunchbase.com/organization/states-title>
- Stribling, D. (2018, June 21). CBRE Puts Machine Learning To The Task Of Figuring Out Good Retail Locations. Retrieved from <https://www.bisnow.com/national/news/technology/cbre-puts-machine-learning-to-the-task-of-figuring-out-good-retail-locations-89846>
- Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer. Retrieved from <https://books.google.com/books?id=bXzAlkODwa8C&printsec=frontcover#v=onepage&q&f=false>

- The Making of Data Science Superheroes at Crest Financial. (2018). *DataRobot*. Retrieved from <https://www.datarobot.com/data-science-superheroes/crest-financial/>
- Thompson, A. K. (MIT). (2018, April 18). Personal Interview.
- Tidman, H., Fischer, M. (Stanford), Ladha, S. (Doxel), & Rippingham, C. (DPR C. (2018, June 15). Construction Under Tech — The Build. *A16z Podcast*. Retrieved from <https://a16z.com/2018/06/15/construction-technology-building/>
- Tidnam, H., Lynn, G. (UCLA), Neff, G. (Oxford U., & Young, T. (PlanGrid). (2018, June 13). Construction Under Tech — Info Flows. *A16z Podcast*. Retrieved from <https://a16z.com/2018/06/13/construction-technology-information-flows/>
- Topos. (n.d.). Retrieved June 19, 2018, from <https://topos.ai/>
- Turner Building Cost Index. (2017), Q4, 3.
- Two Minute Papers. (n.d.). Retrieved June 17, 2018, from <https://www.youtube.com/channel/UCbfYPyITQ-7I4upoX8nvctg>
- Vlaming, W. (RCA). (2018, April 19). Personal Interview.
- Vomero, V. (Foxy AI). (2018, June 7). Personal Interview.
- Wallace, B. (Fifth Wall), & Bruss, N. (2018, May 11). Personal Interview.
- Wang, Y. (Auction. com). (2018, April 30). Personal Interview.
- Wiley, D. (Motionloft). (2018, June 15). Personal Interview.
- Williams, S. (2018). Atlas of Lighting. <https://doi.org/10.1080/00045608.2013.846157>
- Williams, S. (MIT/Envelope City). (2018, June 12). Personal Interview.
- Wong, I. (2017). *Modernizing Real Estate with Data Science // Ian Wong, Opendoor (FirstMark's Data Driven)*. YouTube. Retrieved from <https://www.youtube.com/watch?v=dR5N8cMkIGQ>
- Woyke, E. (2018, February). A smarter smart city. *MIT Technology Review*. Retrieved from <https://www.technologyreview.com/s/610249/a-smarter-smart-city/>
- Zhou, G., Ji, Y., Chen, X., & Zhang, F. (n.d.). Artificial Neural Networks and the Mass Appraisal of Real Estate, 180–188.
- Zillow. (n.d.). Retrieved June 16, 2018, from <https://www.zillow.com/>
- Zillow Prize: Zillow’s Home Value Prediction (Zestimate) | Kaggle. (n.d.). Retrieved May 29, 2018, from <https://www.kaggle.com/c/zillow-prize-1>
- Zirbel, A. (Opendoor). (n.d.). The Stack That Helped Opendoor Buy and Sell Over \$1B in Homes. *StackShare*. Retrieved from [https://stackshare.io/opendoor/the-stack-that-helped-opendoor-buy-and-sell-over-\\$1b-in-homes](https://stackshare.io/opendoor/the-stack-that-helped-opendoor-buy-and-sell-over-$1b-in-homes)