APPLICATIONS OF MACHINE LEARNING IN BUSINESS DECISIONS

JOHN MARCUS III

Applications of Machine Learning in Business Decisions

John Marcus III

www.jwmarcus.com

1st Edition – Summer 2017

Reviews

A great introduction to how the use of Machine Learning to provide context will be the key to boosting business velocity, enabling dynamic business scaling, and ensuring greater business operational flexibility in our new digital world.

Rick Villars
Vice President, Datacenter & Cloud, IDC

John presents business leaders with a clear illustration of machine learning success, from value demonstration to implementation. Having spent a decade in data analysis and customer success leadership, I know what an organized, usable map for business leaders to follow looks like. John has created just that for machine learning with this paper.

Michael Redbord
Vice President, Services and Support, HubSpot

John may be new to the machine learning space, but he is here to stay. With his vision for the future applications of machine learning coupled with his respect and command of the academic underpinnings, John is sure to become a serious force in the space.

Slater Victoroff
CEO, Indico

A great introduction to the idea of machine learning, with a much-needed focus on the importance of good data as the foundation of a useful, intelligent system.

William Carter Huffman
Machine Learning Scientist
Former NASA JPL Technologist

Preface

This document provides a semi-technical primer for executives looking to orient themselves to machine learning and applications of the technology in the process of making business decisions. As the term “Artificial Intelligence” makes its triumphant return as the most hyped buzzword, marketing and sales executives need to separate technological capabilities from promotional fluff. After reading and understanding the material in this guide, you will be able to quantitatively evaluate vendors and partners claimed use of AI according to their mastery of machine learning.

After reading, be sure to join the conversation online about machine learning and decision making. This field continues to evolve and your input helps to drive both the definition and the adoption of machine learning.

I always welcome feedback and discourse, so please reach out to me online at www.jwmarcus.com or via linkedin.com/in/jwmarcus.

Introduction

The emergence of self-teaching systems known as "machine learning" algorithms will cause modern businesses to face a shift in thinking unrivaled since the industrial revolution. In this paper, we will explore both how machine learning might replace human generated models and the devices that will fit this growing demand. We will also discuss challenges associated with this transition from human driven decisions to a business world driven by machines.

Today, machine learning algorithms exist not only in the mathematical realm but also in the workplace. Recent advances in computer hardware now power the immense computations required for creating these algorithms (1). From laboratory models that automatically discover gene sequences to Wall Street's models that seamlessly execute high frequency stock trades, computers now outperform humans in a wide array of simple and complex tasks (2)(3). As we look to stay ahead of the automation, we must understand how this technology will impact not only our business processes but also business decision making, the very essence of our jobs.

As business leaders looking to solve complex problems, we typically leverage our intuition and (sometimes) data to drive important business decisions. However, un-intuitively, our personal preconceptions often limit our effectiveness in making unbiased decisions (4). Entire bodies of science study human cognitive bias and its negative impact on business decision making. Learning models lack this bias and therefore excel with their data-centric predictions. Research shows that properly trained algorithms predict outcomes faster and more accurately than humans in specific data modelling tasks (5). Therefore, we stand to gain immense insight from these computer systems. While predictions help provide insight, the actions and policies that originate from those recommendations drive the actual results.

Currently, machine learning models require humans to implement any suggested course of action. Looking to the future, learning algorithms will not only predict outcomes but also prescribe and execute actions without human intervention. The wisdom offered by these models far exceeds our current utilization of the technology. As such, we stand to make major improvements in decision making by leveraging machine driven intuitions.

According to artificial intelligence analysts, learning models will soon replace many business functions related to prediction. Furthermore, the predictive power of these models will rely not on analyst skill but upon implementation accuracy. These systems will act of their own regard on behalf of the humans they could potentially replace. Given such, we must understand the capabilities and limitations of these new systems to stay competitive and relevant. For example, the capability to accurately predict future events depends on a solid foundation of data.

Machine learning models require substantial amounts of clean and accurate training data to make the predictions (6). Fortunately, the proliferation of big-data creates an abundance of data from which to learn. Unfortunately, acquiring clean and consistent data across multiple systems remains the largest impediment to training models for use in business decisions.

The Current State of Data Management and Decision Making

As executives and leaders, we want to make great decisions and therefore gather data from our normal business activities. We then sit in conference rooms, poring over figures and diagrams in search of insights. Today, business decision best practices focus on model creation and interpretation. However, turning data into insights requires more effort and analysis then we realize. Leaving the obvious problems of inaccurate data aside for a moment, we as humans fall victim to our own internal biases. Leading research on cognitive bias shows that humans make suboptimal decisions for various reasons such as "adverse selection" and "sunk-cost" bias (7)(8). Making data-driven decisions then becomes a balancing act of data collection, data aggregation and model creation.

Until relatively recently, the task of data collection and aggregation contributed heavily to marketing and product development costs (9)(10). The Nielsen Rating audience measurement system provided an example of an ever-changing challenge for sourcing accurate data. Originally based on phone calls to television viewers, this rating system failed to accurately capture viewer engagement in a meaningful manner. Due to the analog, unconnected methods researchers used, they failed to find efficient ways of reaching viewers in meaningful quantities. Even up until the mid-2000's, the proper technology to assess viewer engagement eluded the Nielsen Company, leading to wide criticism of the effectiveness of their methodologies. Furthermore, local television executives made programming decisions based on this flawed data (11). The cost and effort required for accurate data gathering far outweighed the insights derived from the data, leading to the destruction of business value.

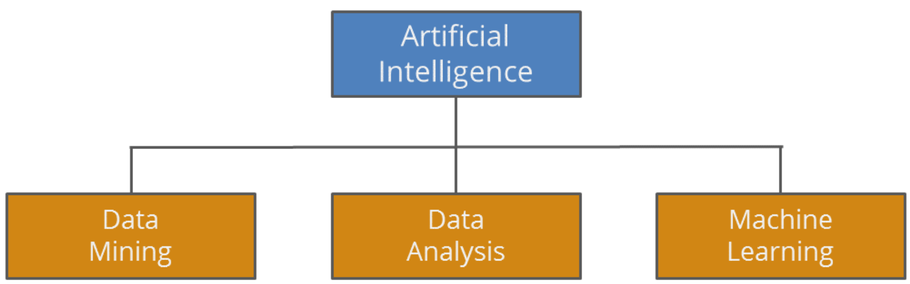
Starting in the early 2000's, advances in digital user tracking and analytics ushered in a plethora of information for use in business planning. Browser cookies, tokens and email tracking pixels all automatically collect information on everything from click behavior to browsing habits. Although business users debate the validity of these data points, the quantity of information continues to grow exponentially. Even with this multitude of data, the problem of aggregation remains a sticking point when making solid business decisions.

Business users need access to their data across all their systems. However, in most cases, only developers possess the tools to extract data from these systems, by directly accessing the underlying databases. Furthermore, according to machine learning experts, only the top 1% of all enterprises utilize API's to export and analyze their own data. This unbalanced access to information precludes business users from using data to help with decision making. Furthermore, most companies lack the engineering talent required to normalize this extracted data into a format useful for comparisons. Because most of the processing happens within a "black-box," many business users hesitate to trust the results presented by data scientists. With inaccessible and inaccurate data, decision makers rely on only their intuition.

The Machine Learning Landscape & Artificial Intelligence

The rise in popularity of any technology drives media and marketing coverage. Artificial intelligence (AI) and machine learning (ML) share this trend but the content used to explain AI and ML varies in quality. The scientists with the most knowledge on these topics tend to write difficult to digest academic papers. In addition, many academics refrain from using AI to refer to ML, seeking to be precise in their definitions. To orient ourselves to this field, we need to understand the vocabulary and capabilities.

Artificial intelligence covers a broad field of both scientific research and practical implementation which strive to replicate "knowledge tasks" traditionally reserved for human thought. By definition, artificial intelligence consists of "any device that perceives its environment." Within this admittedly broad scope, artificial intelligence solves tasks ranging from image detection to shopping cart suggestion to piloting self-driving cars. Think of AI as the umbrella term that covers many distinct fields of knowledge tasks. This umbrella of artificial intelligence contains many disciplines such as data mining, data analytics and machine learning.



Organization Chart of Machine Learning

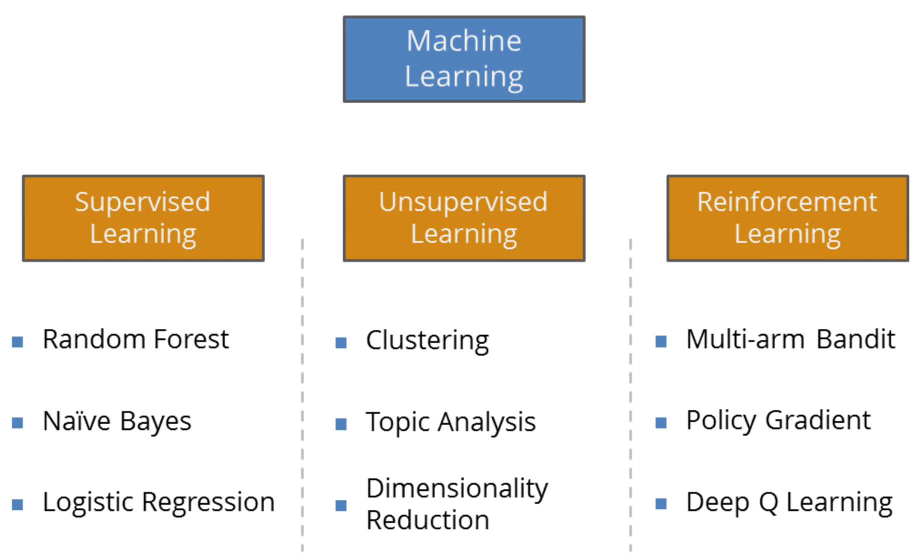
Data mining focuses on the ingestion and relation of data from very large data sets. Engineers and data scientists trying to wrangle data often use specialized software to digest and load data from various sources. These professionals seek to make substantial amounts of data readily available for further investigation.

Data analytics pertains to transforming and modeling data for human consumption. Analysts and consultants regularly use charts and graphs to convey meaningful insights to their audience. These charts and graphs drive decisions that impact every part of the business. As mentioned previously, many times our own human limitations reduce the intrinsic value of data analytics. Yet we continue to use data analytics heavily in our corporate lives to help guide our understanding of the business.

Machine learning describes a third class of problem-solving via the use of self-enhancing computer programs. The programs use software in a continuous improvement process like human learning. For humans, our experiences shape our ability to learn and make predictions (12). While learning new skills, we experience events and adjust our mental models based on the correctness or incorrectness of those actions. Similarly, machine learning uses advanced mathematics to simulate the feedback and adjustment process within computer models. Some architectures go as far as to network models that create "layers" of learning (13). Most promisingly, these systems deduce patterns and predictions automatically, without explicit instructions on how to do such.

Types of Machine Learning and Their Applications

Experts most often describe machine learning in terms of the three major types of algorithms used to solve problems; supervised learning, unsupervised learning and reinforcement leaning.



Types of Machine Learning Algorithms

Supervised learning algorithms start with a problem, then take properly collected and categorized data, called training data, to make a guess as to a solution. After making a guess, the algorithm next learns the answer to the question. Based on its performance, the model updates its own parameters to improve future predictions. This process repeats millions, if not billions of times to slowly improve the intelligence of the machine learning algorithm. Note that programmers do not tell the algorithm how to evolve at each iteration, but the math involved dictates how the model learns over time. Supervised learning requires properly labeled, previously existing data to learn. Because of such, these methods predict outcomes from previously labeled data to make predictions into future results (14).

For an example of supervised learning, let us try to predict the marketing persona for new leads in a marketing automation system. Specifically, we wish to tag new leads to one of four different (previously identified) buyer personas. First, we categorize existing leads such that all training data belongs to a buyer persona. Next, we train a *logistic regression* algorithm using the marketing and buyer persona data from these leads. Finally, we pass a new lead to the trained model, which gives us probabilities as to which persona this new lead might belong to. The largest value indicates the most probable buyer persona for that new lead, based solely on their marketing data. We can now use this persona information to trigger new campaign actions like automated emails.

As another example: if we want to predict the estimated value for a new sale opportunity, we might utilize *linear regression* to provide insight. Using historical opportunity data matched with closed account information, this model learns to predict an "expected dollar value" of a closed deal.

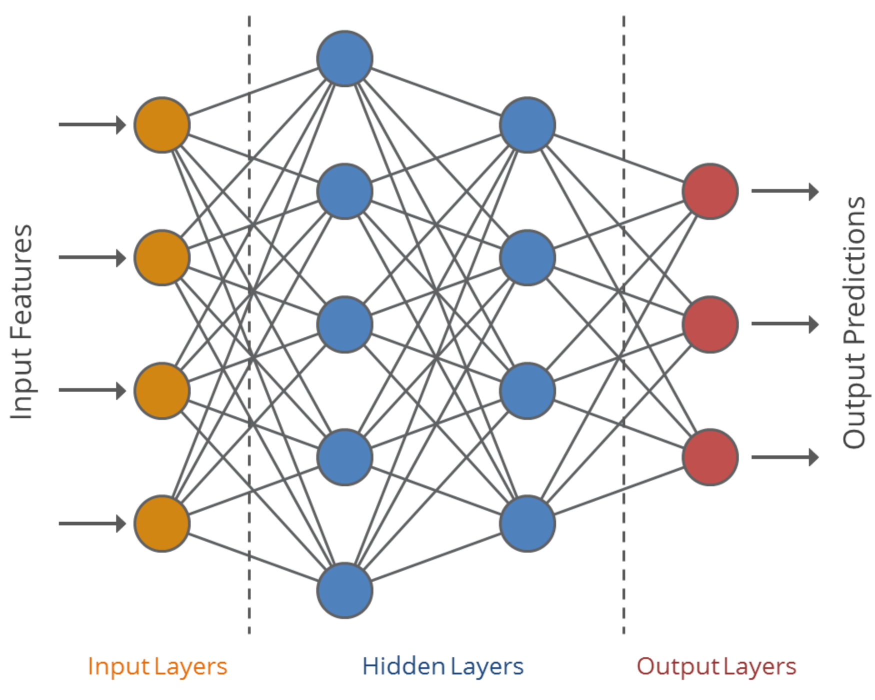
Unsupervised learning algorithms do not require labeled data to provide insights. Instead, these models tend toward either reducing complexity or clustering data. Lacking labels, these algorithms create their own hypotheses by measuring the similarities and differences between the data.

Using the prior example, we can take an unlabeled marketing database and create marketing personas. By using an unsupervised algorithm such as *DBSCAN*, we can organize our existing data into distinct groups based on non-obvious similarities. After clustering, a marketing team could identify and label each persona with some meaningful name. Once we add a label to the data we transition from unsupervised to supervised learning once again.

Reinforcement learning defines the third type of machine learning category, whereby an algorithm takes a target goal and a scoring system by which to measure its progress. This category of learning relates well to problems related to short term trade-offs.

Inspired by behaviorist psychology, reinforcement models use software agents to solve difficult problems via goal seeking behavior. These algorithms look for optimal behavior by taking actions toward achieving a goal and adjusting to further improve performance. The methods use a reward function to provide feedback to the algorithm and the model further changes to maximize that reward function. An example model might seek to maximize profit when evaluating which of a dozen sales opportunities to pursue.

In addition to these three most common modes of learning, some shared terminology comes up when researching machine learning. The term *neural network* refers to a family of loosely related techniques used in all types of machine learning. This design acts as a blueprint for solving difficult problems across all artificial intelligence.



Neural Network Model with Hidden Layers

Neural networks, while not a learning methodology on its own, appear frequently in the use of machine learning. These networks connect sequences of logic "units" in layers which feed into further layers. Each unit contains parameters, called weights, which determine what the unit will output given various inputs. Each neural network consists of an input layer, an output layer, and one or more "hidden" layers between the input and output. Furthermore, *deep neural networks* simply contain two or more hidden layers.

These various methods of machine learning require sufficient quantities of computing power and data to make accurate predictions. Until recently, the hardware required for the underlying mathematical operations in advanced machine learning simply did not exist. With advances in technology, data scientists now balance the need for hardware and data in their implementation of machine learning.

Hardware, data and implementation

The mathematical formulae for machine learning date back decades, but computers could not run these calculations in a reasonable amount of time. Interestingly, the increase in computing capability came from advances in the video gaming hardware. To run graphic intensive video games, gaming PCs and gaming consoles leverage Graphical Processing Units (GPUs). GPUs execute mathematical operations differently from a normal computer processor, using highly parallel processing which makes this hardware ideal for running specific algorithms. The adoption of these specialized processors now drives down costs and funds further research into GPU hardware specific to solving machine learning algorithms (15).

Regardless of the algorithm chosen for the task, all machine learning models need data to operate and evolve. Given inconsistent and erroneous data, the algorithms will fail to adapt to the underlying structure and generate unhelpful models. Consistent and normalized data makes or breaks a great machine learning implementation (16).

For this reason, business consultants and operations experts seek out unified sources of data with which to train their learning models. When working with business systems, data scientists struggle to create a consistent view of the data across all business departments. Various tools available today offer help exporting data from systems, but most all of them require skilled data mining scientists to reconcile the information between the systems.

Other tools purport to synchronize business data but do not reconcile and normalize information across the synched objects. Due to the complex nature of many solutions available on the market, data preparation continues to be a major obstacle for many machine learning projects.

Even with cleaned and normalized data, machine learning algorithms tackle specific subsets of business problems. Today, algorithms primarily work on tasks related to prediction and clustering, requiring skilled data science engineers to pick the right tool from many available options. Multiple algorithms solve similar problems so even with excellent domain experience an engineer could still choose the wrong method. The extensive effort required to build purpose-specific models leads some businesses not to invest in machine learning technology.

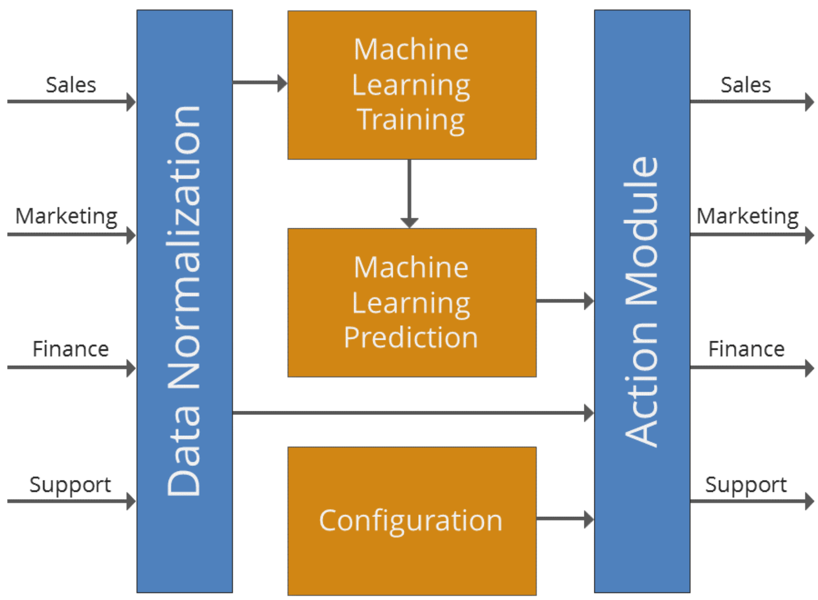
As domain-specific models improve, scientists combine different systems together to create more complex learning models. For example, unsupervised algorithms (such as *dimensionality reduction*) can feed into supervised algorithms (such as *logistic regression*) to form more robust learning systems. This evolution continues to drive toward a truly general-purpose version of artificial intelligence, called *artificial general intelligence (AGI)*. Most experts estimate that AGI will take decades to create (17)(18)(19). Until then, we need to pick the right tool for the job.

Regardless of how scientists construct the models, humans still need to drive the adoption and execution of the recommendations. If the stakeholders do not trust the data or the model, they tend not to directly implement the suggestions from the data models. We as end-users modify the recommendations generated by the models to fit our personal beliefs and view of reality (20). Introducing human bias reduces the value of the learnings generating by the models and the potential benefits they could provide. To avoid this flaw, data scientists must not only analyze the data but also predict a course of action.

Future Business Applications of Machine Learning

Within the next two years, we will see prevalent use of machine learning models that both create action plans and automatically execute those actions to drive business growth. By acting without human intervention, the models will avoid the limitations on effectiveness that we impose through our own biases.

Let us define a hypothetical "machine learning decision module (MLDM)" which combines both machine learning and data automation into one independent system. Next, let us take any business which sells software as a subscription service. This business wants to predict future customer cancellations to mitigate the cancellation rate (churn rate) of their existing clients. When an analysis of the data predicts a customer to be a churn risk, we want the MLDM to direct an account manager to reach out to the client and prevent the churn proactively.



Model for Machine Learning Decision Module

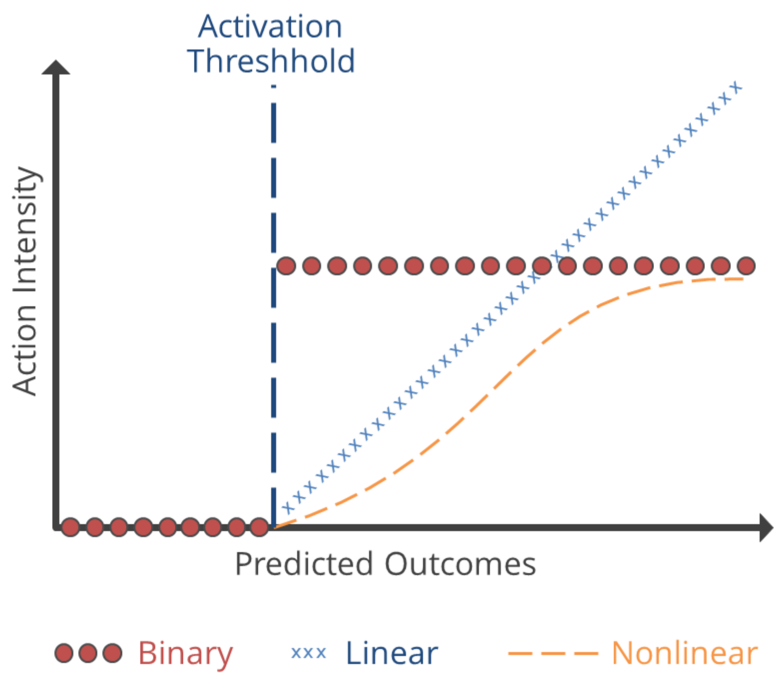
To accomplish this, we first train a MLDM to predict a cancellation based on several factors like account age, product engagement and number of help-desk tickets opened in the past few months. In addition, we give the module a set of corrective actions to execute in the case of a predicted churn. When the MLDM predicts a customer at risk of churning, it executes a sequence of tasks to minimize the potential risk to the recurring revenues. In addition, by leveraging *online learning* models, the module consumes real-time results to retrain and improve the model further as new data becomes available (21).

As the module operates independently from humans, it side-steps issues of cognitive bias found in human implemented systems. When put into practice, Machine Learning Decision Modules operate in three primary modes; binary action, linear action and non-linear action.

Binary action modules make predictions about a future state and take actions when the predicted outcome crosses a predefined threshold. Like a smoke alarm blaring at the first sign of a fire, remediation actions trigger immediately with no ramp involved. The modules may stop when the predicted outcome no longer crosses the threshold. This configuration excels working on problems that require on or off correction such as error detection and logging.

Linear action modules take actions in progressively larger steps as the predicted outcome from the model progresses further past the activation threshold. This mode of operation takes the expected state of output and applies ever-increasing remediation to match and control the projected output. A throughput control valve on a water dam could represent a linear action model. When the water level rises, the dam lets more water through to keep the water level consistent. As more rain falls and water levels rise, the dam opens the flood gates further. As waters recede, the controls restrict the gates to their nominal levels. Depending on configuration, this action mode could reinforce conditions and accelerate potential positive effects as well. These modules work well when tightly managing an element of a system to within a defined tolerance.

Nonlinear action modules take actions in varying levels depending on how far from baseline the model predicts that outcome. In contrast to linear action modules, nonlinear modules eventually reach a saturation point or maximum action point. As an example, a doctor may prescribe pain medication to a patient after surgery. When the patient reports more discomfort, the doctor administers more medicine. However, the doctor never allows more than the maximum safe dosage to the patient, regardless of the pain reported. The nonlinear action module works best with fixed available resources.



Decision Module Action Modes

Realistically, machine learning decision modules (MLDMs) would trigger any of a multitude of actions depending on the business goals. The limits of these decision modules depend on the interconnectivity of the systems involved. If a MLDM connects to a sales system, it gains the ability to drive sales actions by creating open tasks for the sales team. Alternatively, connecting a MLDM to a marketing automation system will grant the module the ability to enroll leads into email campaigns.

Purpose built decision models require both training data and heavy computing power. We discussed the challenges related to clean data earlier. For decision modules of the future, the most effective devices will combine not only the software required to derive decisions but also the hardware needed to compute the mathematics. Manufacturers may build specialized devices and task specific appliances for businesses to buy, configure and use.

Turning over a business decision-making process to a machine raises issues of trust. However, a clear understanding of how the models derive their predictions helps the business to come around to trusting the actions taken by the machine learning decision modules. Businesses need a corporate culture driven by data for the best chance for success. Furthermore, if the stake holders do not act on the recommendations then they will lose out on any potential benefits that the models provide.

To trust and embrace machine learning quickly, we must understand the limitations of machine learning via both in terms of the data integrity requirements as well as the potential limitations to data access.

Data Caveats and Privacy Concerns

A more expressive set of data often leads to better trained models (22). The challenge remains in reconciling data between multiple systems. Businesses naturally segregate their internal data into silos based on the divisional divides within the company. For example, marketing data rarely reflects perfectly into the sales or finance systems and vice-versa.

When leveraging data from multiple systems, problems arise when matching data across business units. For example, a contact in the marketing automation system correlates to a contact in the customer relationship management system. In this case, the *email address* field ties the two systems together. Relating this contact to a product or order in the finance system poses a significantly more difficult challenge.

Companies may invest in integration technology to move data back and forth between systems, but in most cases this information remains in disparate data stores. Furthermore, unless the unified information resides in a centralized data structure, machine learning algorithms will find data ingestion difficult. By making centralized, integrated data across business functions a priority, companies ensure their ability to leverage the best that machine learning offers today and in the future.

Data privacy in the use of machine learning remains an unexplored realm of both legal and ethical concern. The release of this innovative technology raises the question of how companies and algorithms learn from our data. One grey area in the current law relates to how scientists use confidential data to train these models.

Data scientists lack the ability to reverse engineer a trained machine learning model to reconstruct the original data. This provides some level of comfort to those willing to offer their data for use. However, the effectiveness of the model relies on the learnings gained from the training data. In some sense, the very existence of the model requires absorbing some "essence" from the data. The question of who retains the rights to the learnings from our data remains unresolved.

Wrap-up

In this paper, we explored the concept of machine learning and discussed a potential application to replace the human decision making process. Soon, decision modules will decide and act on our behalf, optimizing our business automatically and better than ever. To be clear, preparing to act requires an investment in learning. Machine learning already shapes our decision making today and businesses will adopt future versions of this technology faster than expected. We must invest in understanding the landscape and applications of machine learning today to get ahead of this fundamental change in the way we do business.

About the Author:

John Marcus III is Entrepreneur in Residence at .406 Ventures. Previously he co-founded Bedrock Data and served as CEO of the business during its founding and growth phases. Prior to Bedrock Data, John held multiple roles at HubSpot (NYSE: HUBS). As a serial entrepreneur, John has a passion for making novel technology simpler to use and sharing those best practices across marketing, sales, support and technology communities.

Sources

Referenced Sources

1. http://datascience.ibm.com/blog/the-mathematics-of-machine-learning/
2. https://dupress.deloitte.com/dup-us-en/deloitte-review/issue-20/augmented-intelligence-human-computer-collaboration.html
3. http://www.theverge.com/2016/4/28/11526436/mark-zuckerberg-facebook-earnings-artificial-intelligence-future
4. https://www.entrepreneur.com/article/279927
5. http://www.datasciencecentral.com/profiles/blogs/human-brain-vs-machine-learning-a-lost-battle
6. http://www.ntu.edu.sg/home/egbhuang/pdf/ELM-Clustering.pdf
7. http://www.sciencedirect.com/science/article/pii/S0263786316302769
8. http://www.hbs.edu/faculty/Publication%20Files/08-102.pdf
9. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4391295/
10. http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-124
11. http://www.pbs.org/wgbh/nova/listseason/19.html
12. http://scienceblogs.com/gregladen/2014/03/07/is-human-behavior-genetic-or-learned/
13. http://www.annualreviews.org/doi/abs/10.1146/annurev-vision-082114-035447
14. https://pdfs.semanticscholar.org/d119/50567411a4c11b7e0a4ee72e6b4a2e182aed.pdf
15. https://www.gamespot.com/articles/the-past-present-and-future-of-the-gpu-according-t/1100-6421893/
16. https://arxiv.org/pdf/1703.01780.pdf
17. http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html
18. https://techcrunch.com/2016/12/14/why-we-are-still-light-years-away-from-full-artificial-intelligence/
19. https://blogs.microsoft.com/next/2015/07/06/the-future-of-artificial-intelligence-myths-realities-and-aspirations
20. http://www.aaai.org/ojs/index.php/aimagazine/article/view/2513/2456
21. http://www.jmlr.org/papers/volume11/chechik10a/chechik10a.pdf
22. http://web.mit.edu/vondrick/bigdata.pdf