

Analytics Evolution Guide

Enterprise and Cloud IT
Artificial Intelligence and Machine Learning



Moving Up to Artificial Intelligence and Machine Learning

Gain Business Value Using Machine Learning and Artificial Intelligence



Esther Baldwin

Solutions Architect, Intel

Dr. Steve Chadwick

Sr. Principal Engineer, Intel IT

Zeeshan ul-hassan Usmani

Subject Matter Expert, Consultant

Jim Fister

Principal, The Decision Place

The Scope of Business

The fundamentals of a successful business have remained consistent for decades, if not centuries. A good business knows its customer and provides the best product ahead of its competition. Decades ago, a shopkeeper learned the customers' names, knew what they bought and when they needed it, and did the extra things to ensure the individual that he or she was valued by the shop. The owner of a bookstore knew to stock material based on bestseller lists and critical review with some additional insight from the people who frequented the store. A good hardware store manager would monitor the change of seasons and the change of stock, while finding ways to reduce inventory with special sales that attracted the buyer.

While these business fundamentals have not changed, the pace and scale of business have created a whole new set of demands on a company. A modern business has access to a much larger set of customers with a much wider array of desires and needs. However, worldwide competition has access to the same set of customers which pushes a business to respond faster. Business at the speed of the internet requires that the transaction be presented, considered, and closed before the customer decides to look elsewhere for a better option. The greater rate and volume of messages presented to a customer means that any individual message must be customized to have maximum impact to break through the noise.

Insight itself has also changed. While business in the past could rely on a few sources of trusted data to make decisions, the validity of today's information must be constantly tested. The massive increase in data available to a

Table of Contents

Contents

The Scope of Business	1
Business: More than Data	3
Beyond the Traditional Database	3
Intel and Machine Learning: Fault-detection Control Using Machine Learning	4
Developing Modern Analytics Solutions	5
Success Requires Widespread Change	5
Business Intelligence Today	6
Intel Products Provide the Foundation	7
A Spectrum Across the Entire Maturity Model	8
Conclusion	9
Appendix: A Sampling of Tools for AI and ML	10

business -- both within and outside its corporate control -- requires verification to properly apply it.

Therefore, a business needs every advantage available to it in this environment. This puts pressure not only on the leading business unit, but also on the IT department that supports the business. The pace of innovation requires the company to

respond with new systems and processes to stay ahead of the

competition. The opportunity for the IT department is to also respond as a business itself, becoming a partner

rather than a supplier to the

business unit. This allows IT to understand the business and the problem in depth, and that understanding helps the business with optimal solutions.

Leaders will often agree that managing change is their biggest challenge.

People, and organizations, resist change. However, there is a simple framework that can be utilized to set the organization on its way to success:

- Managing Information – both scale and scope of the data
- Funding Analytics – R&D budget allocated to discovery
- Building the Capability – moving to operational deployment
- Assessing the Business Value – re-investing where ROI is high

The relationship between the process and the tools is indirect for most companies. Many tools can be used to deliver on simple data evaluation, but that will not move the company forward. A company needs to understand that creating and adopting a strategy for getting the most out of data can be accomplished by building the IT/business partnerships, as well as following basic methodologies to success.

New tools in the areas of artificial intelligence (AI) and machine learning (ML) are emerging to help businesses succeed at all levels of maturity.

Understanding how they apply must first come with an understanding of how businesses use data today and how that can evolve.

Being successful using advanced analytics tools is as much an organizational opportunity as it is a technological one.

Business: More than Data

It's not enough to just gather data. Unless your company can add context to all the data you have to support your decisions, the data alone might not be useful for the purposes of your corporate strategy. Creating a context for your data, however, allows it to narrate a story that builds to a conclusion for a successful business.

In a modern IT department, data analytics is already playing a key role and becoming a more useful tool for your company. Since a good business runs on data with the proper context, the IT organization must constantly look for new methods to acquire, store, analyze, and enable business decisions that act on the data in a timely fashion. This is the difference between winning and losing in today's fast-paced environment.

This is not a one-time effort, however. The pace of innovation in data analytics has been increasing as fast as the growth of data, if not faster. New tools are available at every turn, and that means that new skills are required in the IT department to provide that necessary context. The churn also requires a good knowledge of the tools and underlying systems that support them, as well as an understanding of ways to store and format the data that allows new tools to be immediately applied. This likely translates to new systems, new personnel, and an ever-increasing storehouse of contextual data. All of this might mean higher cost of operations that demand justification for your business.

There is a delicate balance between taking on costs in

If you woke up as an industrial company today, you will wake up as a software and analytics company tomorrow.

—Jeff Immelt, CEO, General Electric*

terms of new people, new tools, and the ROI gains that offset

these costs. But -- like any good IT process -- the value gained for both operations as well as business insight will be necessary, especially knowing that the competition has access to the same opportunities as your business does.

Beyond the Traditional Database

Business analytics has traditionally been an extension of the database that runs the company. And as such, it has relied on many traditional database methodologies to gain insight. Performance gains have come through technology innovation that has essentially kept the data formats and methodologies the same. For example, recent advances in data compression paired with fast solid-state disk and large memory technologies have led to in-memory databases that are significantly changing the heart of the enterprise. Now though, cloud-based and unstructured data technologies have led to new methodologies in data processing.

All of these advancements, however, have exacerbated

Tools that utilize AI are available at many levels of the analytics stack, and can simplify your path to productivity.

existing skills gaps, as well as created new ones. Each advancement of analytics technology means new learning, but also means a new tool that must be understood to truly harvest the results.

While artificial intelligence and machine learning tools are not necessarily new, their recent application to data analytics for business is starting to change the way analytics will best be done. Recent efforts to incorporate more standard tools using these methodologies have made rapid advancements in the way businesses make decisions. New technologies are emerging regularly,

Advanced Analytics Move-Up Guide

and these provide new opportunity for any company willing to test them. More importantly, these new tools are not necessarily ends in themselves. Rather, they are better vehicles to a decision-making destination. In some ways, these tools have similar downsides to standard analytics tools. They do require new skills and new learning, and the time from implementation to value

can still be long until these skills are mastered. However, the return on business investment over time will be incremental immediately in many cases, and the long-term ROI accelerates rapidly, potentially providing a significant increase in insight without a similar increase in human capital. As such, these tools offer a promise of growing the business faster while keeping costs lower.

Intel and Machine Learning: Fault-detection Control Using Machine Learning

At a typical Intel factory, there are thousands of sensors collecting massive amounts of data every day; the resulting data sets are huge. In fact, Intel generates over five billion data points per day per factory, and this is growing with each technology node. Over time, Intel has allocated resources and tools to help manage and harvest the value from this data. For those companies getting started, you should realize that the potential of data doesn't require a vast network of sensors in the beginning; it starts with understanding how to manage and optimize whatever sensor data is available. Starting small allows for a very focused effort on a specific, tangible problem. In the early stages, the Intel team experienced a learning curve on how to mine and integrate data across multiple sources. Over time, they were able to identify what was most important in the vast sea of information and began to discover actionable insights from the data. The algorithm they developed to assist them started with a single variable which they slowly began adding complexity to. Today, the analytics our analytics activities have grown to support increasing amounts data, in addition to more traditional sources. The company is confident that, as others undertake this journey, they will also realize the benefits they can obtain from their own data.

Key Outcome: *Ability to process over five billion data points per day per factory, resulting in measurable improvement in equipment availability and yield*

Methodology: Many of Intel's factory equipment environment sensors collect hundreds of data points per second. They extract this data and use it for both real-time detection and end-of-line correlation. Engineers and manufacturing technicians are able to use the insights reported to them to refine equipment behavior.

Because sensor data can often be large and complex, Intel considers the following specific questions when designing its internal storage systems and algorithms:

- How will the data be consumed?
- How is the data stored (schema)?
- Where does the data reside?
- How fast can it be retrieved (this affects hardware choices)?

It is important to start with something simple and build from there. The time series analysis, multivariate analysis and machine learning Intel needs to do on a regular basis is quite complex, but our first attempts began with only a single variable. It's important to take the first step and discover the measurable results, and your ROI can increase rapidly once the organizational momentum builds.

Developing Modern Analytics Solutions

The most effective analytics solutions start with the development of a business case. This is best when done in conjunction with the business unit that will act on the results. As the IT organization moves from supplier to partner, it will establish its credibility in the business solution space for your company.

Establishing the data available and building the contextual framework is the next step. The business unit can identify the existing data sources that are used as well as additional data it has that could be added. At this point, it's useful to seek additional data that exists inside and outside the corporate walls. This search can, and should, be far-ranging -- from expert industry data,

Plug-and-play tools can match the processes that an IT department is already using, and the tools are increasing in variety and usefulness.

historical patterns that might apply, seasonal adjustments, and even social media. The data sources can then be

studied for pattern, format, and frequency of update. This gives the organization options for how to train its systems and analyze current and future input. It's critically important, especially in deployments that utilize artificial intelligence, to understand the validity of the data that could be used. Many traditional deployment methodologies will have a pre-processing phase that can eliminate outlying data. Learning-based tools will usually incorporate all data available in order to glean information from the variability. However, if the data is genuinely bad, or has significant errors, it can teach the system to take the wrong direction.. As the skill of the IT organization to use the tools increases, the system can be trained to recognize the bad data.

In general, artificial intelligence and machine learning tools utilize iterative operations that examine all the data in its various forms to look for patterns. With additional data, the tool will continue to learn and generate more reliable outcomes. While some of these operations will require time to be effective, the ROI is likely to accelerate with organizational experience.

A well-functioning IT organization can use the opportunity to develop skills around a common set of new tools that can be reapplied to various business solutions. Immediate opportunity can be found by looking for problems that have similar data sources and forms.

Success Requires Widespread Change

It's important to understand that any effort to adopt modern analytics as a way to do business is as much or more a change in organizational will as it is a change in tools and processes. The company has to undertake a commitment to use information in new ways and to change its business outcomes on the generated recommendations. There's a natural resistance at a management level to simply, "follow directions after pushing a button." Strong leadership is necessary to make the changes.

The IT organization must also move away from just being, "ticket-fillers," that solve the problems given to them. Business and IT leadership need to agree to involve the subject experts and define business requirements. Owners of data need to be consulted to free data from internal restrictions. The overall organization must be filled with willing and enthusiastic partners that identify problems and the data that can be used to solve them.

All these new skills mean positive change. This change makes the transfer of the human capital to new opportunity easier, making the IT organization, and the company overall, more nimble.



Consumer

Smart Assistants
Chatbots
Search
Personalization
Augmented Reality
Robots

Health

Enhanced Diagnostics
Drug Discovery
Patient Care
Research
Sensory Aids

Finance

Algorithmic Trading
Fraud Detection
Research
Personal Finance
Risk Mitigation

Retail

Support
Experience
Marketing
Merchandising
Loyalty
Supply Chain
Security

Government

Defense
Data Insights
Safety & Security
Resident Engagement
Smarter Cities

Energy

Oil & Gas Exploration
Smart Grid
Operational Improvement
Conservation

Transport

In-Vehicle Experience
Automated Driving
Aerospace
Shipping
Search & Rescue

Industrial

Factory Automation
Predictive Maintenance
Precision Agriculture
Field Automation

Other

Advertising
Education
Gaming
Professional & IT Services
Telco/Media
Sports

Business Intelligence Today

It's easy to find the work and promise of artificial intelligence in daily life: from virtual assistants to voice recognition, from personalized recommendations on movies and books to nascent self-driving cars; and from yearly sales forecasts to stock market predictions. As seen in the figure above, the possibilities for business span every major industry. These implementations are modernizing IT architecture and providing increased ROI to the company in many cases.

Here are a few short examples of artificial intelligence being used by the industry in innovative ways:

- Using AI to Build New Market Segmentation:** Muslims make up 24% or 1.8 billion of world's population. They spend an equivalent of US \$322 Billion worldwide on clothing and footwear annually, a figure expected to reach to \$500 billion by 2020. A UK-based startup, Islamic Design House* (IDH*), wants to capture this market segment by offering Islamic fashion using artificial intelligence and big data methodologies. The company uses unsupervised learning for customers' clustering, and regression analysis algorithms for behavioral segmentation of the online visitors to

their site. Machine learning helps them to define their target market segments and reach their customers with personalized contents.

AI gives IDH the power to understand, reach and influence 780 million young Muslims across 50+ nations to promote the brand and create a new market for Islamic Fashion. By using AI tools, IDH grew from one country to eleven and increased its sales significantly in less than three years.

- Image Recognition and Processing:** JD.Com* is one of the largest online retailers in China with approximately 25 million registered users. The company needed a better solution for processing and categorizing images in its massive database. After building a highly-optimized image extraction and recognition pipeline, the company used Apache Spark with Intel Big DL to build image recognition and feature extraction models.
- Game Theory and Gambling:** Computer science researchers from Carnegie Mellon University* have developed an AI tool called Libratus* that has defeated professional poker players in one of the most difficult forms of the

play – No-Limit Texas Hold'em Poker. You can read the complete paper [here](#) to find out how to tackle bluffing in poker and all other situations with imperfect information.

- **Intelligent Farming:** a Singapore-based company [ADDO.AI](#)* is helping farmers by providing timely recommendations on planting, crop management and harvesting through their hyper-spectral satellite imaging powered by advanced machine learning algorithms to process past yield data, weather patterns and environment mapping.
- **Financial Pattern Recognition:** A leading global finance company took on the challenge to identify known patterns or anomalies in market trading data in order to fraudulent activity or spikes in trading volume. By using Intel® FPGAs designed in an AI framework, the company ingested a large amount of public order book data to automatically learn informative patterns of activity in the historical data.
- **Teaching English:** a Taiwan based startup [ELSA](#)* – English Language Speech Assistant* is mining native language spoken patterns to guide its users on how to advance English language skills and better their pronunciation.
- **Heart Condition Diagnosis:** Mt. Sinai* hospital is using AI to examine 10,000 patient attributes that helps to determine the difference between a fairly common and treatable heart condition and a similar but much more severe condition. The system aids cardiologists and greatly increases accurate diagnosis.
- **Genomic Data Analysis:** The Broad Institute* is a collaboration building a fully open-source Genomics Analytics Tool-kit* (GATK*) to build, optimize, and widely share new tools and

infrastructure that will help scientists integrate and process genomic data using machine learning techniques.

- **Rural-Urban City Planning:** [Zillpo](#)* holds a large array of data points about small villages, towns and cities. It analyzes this data to help government authorities to forecast the population demand on land, traffic, transport infrastructure, and jobs market.
- **Matchmaking for Call Centers:** [Afinity](#)* uses artificial intelligence and big data to pair customers with call center agents based on interpersonal behavior mapping. For example: a caucasian widow with three children living in the Southeast US would be matched to a customer service agent with similar profile, which helps increase the overall chances of successful call.

Intel Products Provide the Foundation

The success of any computing solution, especially a scalable analytics deployment, should be built on a firm foundation. In the past, a solution built for business analytics could be modified and patched with side-car additions. However, the performance impacts of multiple data moves are proving to be too costly. A good solution in the modern business environment is immediately applicable, and significantly scalable over time to address the needs of the business without significant design additions.

Scalable hardware platforms based on Intel® Xeon® Scalable processors provide a powerful and versatile computing baseline that can adapt quickly to new software and algorithms that improve your AI-based analytics. It's not enough, though, to just have a strong compute component. The volume of data in many solutions requires a balanced system with an efficient network topology as well as low-latency storage to feed the compute elements. Ensuring that the deployment

Advanced Analytics Move-Up Guide

can flexibly support the latest networking improvements enables both system performance as well as future scalability.

For demanding solutions that require specialized processing, Intel offers AI-specialized products, the Intel® Nervana™ Neural Network Processors (NNP). These devices utilize a high-bandwidth memory solution to provide the maximum level of on-chip storage and utilize separate pipelines for computation and data management. The architecture is highly parallel and designed with interconnect technologies that ensure maximum scalability for neural network processing. The advent of solid-state disk technology has greatly improved storage solutions in server deployments for nearly a decade. Recent advances, such as Intel® Optane™ Solid State Drives (SSDs), have increased overall storage capacity per system compared to prior SSD products while continuing to reduce latency of data access. When properly backed by a “warm” store, a flexible storage solution will ensure best performance of artificial intelligence and machine learning.

Data movement is also incredibly important, both to get data to an analytics unit for processing as well as within the cluster itself. Intel offers the Intel® Omni-Path Architecture (OPA) fabric solutions ensure high-bandwidth and low-latency connectivity within a scalable system that ensures consistency of parallel processing and data access. Additionally, Intel and others provide a variety of Ethernet* solutions that can move data across standard networks.

For developers, Intel continually provides improvements in performance for the latest hardware via an extensive set of performance primitives and libraries. This ensures that any new system can take immediate advantage of performance without significant rework in deployed software. Many of these libraries are also available in open source, which enables quick adoption of improvements.

By delivering a systems-level approach combining AI hardware, enabling software and tools, memory and storage, and interconnect, Intel supplies the complete foundation for scalable analytics solutions. In addition, Intel shares its best practices via the [IT Peer Network](#), enabling expertise to be easily shared.

To see the full spectrum of Intel solutions for artificial intelligence and machine learning, visit: <https://ai.intel.com/>.

A Spectrum Across the Entire Maturity Model

No matter where your company currently sits in its maturity on data analytics, there is opportunity to add artificial intelligence and machine learning to your portfolio. Almost every major-market analytics package from mainline vendors supports artificial intelligence through add-on libraries. Consult your analytics vendor for options that might already exist in your environment. Also, many common tools that are likely already in use in your company support a variety of plug-in libraries that enable direct access to artificial intelligence. These are also likely optimized for Intel® architecture to maximize performance. There are dozens of learning tools available in the market that businesses can use to get started using machine learning or artificial intelligence for their commercial interests and easily upgrade their capabilities.

Intel also has a variety of open-source tools and frameworks that have been optimized for best performance on Intel architecture. The appendix to this document offers a sampling of different software. More updated information can be found at: <https://software.intel.com/en-us/ai-academy/frameworks> or <https://ai.intel.com/technology/>

Conclusion

It doesn't matter whether your company is just getting started in machine learning and artificial intelligence, or if it's already highly experienced and looking to grow your analytics solutions to a new level. The opportunity to utilize the new solutions that are emerging every day can grow your business. It's key to remember just a few things to keep your journey going:

- Being successful using advanced analytics tools and processes is as much an organizational opportunity as it is a technological one, you have to give teams time to learn.

- Learning is a vehicle, not a destination - Tools that utilize AI are available at many levels of the analytics stack, and can simplify your path to productivity.
- AI has many advantages for organizations that do not have deep analytics background or experience, ML using non-traditional statistics takes more skills, but is immediately applicable to your data.
- Plug-and-play tools match the methods and processes that a successful IT department is already doing, and they are increasing in variety and usefulness.

"In a way, AI is both closer and farther off than we imagine. AI is closer to being able to do more powerful things than most people expect -- driving cars, curing diseases, discovering planets, understanding media. Those will each have a great impact on the world, but we're still figuring out what real intelligence is."

— Mark Zuckerberg, CEO, Facebook*

Appendix: A Sampling of Tools for AI and ML

Tools for Artificial Intelligence and Machine Learning are expanding nearly every day. While this is not an exhaustive list of tools, here are few options to get your company started. Many of these are likely in your environment today.

Solution/Tool	Description	Possible/Likely Uses
Hadoop* and Spark* with MLib*	At the heart of the big data movement, Apache* Hadoop and Apache Spark use the efficiency of Intel® architecture to power base data analytics. The MLib library enables current users of the libraries to add machine learning to better discover patterns and new insights. This library is optimized for Intel architecture	<ul style="list-style-type: none"> • Pattern matching • Classification • Recommendation
BigDL	Distributed deep learning library and learning framework for Apache Spark. Add deep learning functionality to your big data workflows by writing deep learning applications as standard Spark programs. This framework is optimized for Intel® Architecture	<ul style="list-style-type: none"> • Training • Recommendation
R*-Programming	R is a software programming language and environment for statistical computation and graphics. It is written in C*, FORTRAN* and R itself. It's been widely used by data science community for in-depth statistical processing, data pre-processing/cleaning, and data classification. There are more than 11,000 ready to use R packages available that makes it the language of choice when it comes to data. Many of these packages are optimized for Intel® architecture. R is one of the most widely-used open-source programming languages for data science developers.	<ul style="list-style-type: none"> • Linear and non-linear modeling • Time-series pattern analysis • Classical statistics • Supervised and unsupervised learning • Collaborative filtering • Exploratory analysis
TensorFlow*	TensorFlow is an open source software library for high performance numerical computation, highly-optimized for Intel architecture. Its flexible architecture allows deployment from single servers to clusters of servers to edge devices. Originally developed by researchers and engineers from the Google's* AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains.	<ul style="list-style-type: none"> • Computational analytics • Statistical processing • Classification
MXNet*	MXNet is a highly scalable, lean, and flexible deep learning framework to train and deploy neural networks. It provides a portable library that can scale to multiple machines. This deep learning library also supports convolution networks (CNNs) and long short-term memory networks (LSTMs). It supports both imperative and symbolic program languages. It works with a wide variety of programming languages, and is optimized for Intel architecture.	<ul style="list-style-type: none"> • Image/object recognition • Feature detection • Character recognition • Multiple classification
H2O.ai*	H2O.ai is an open source, fast, in-memory, distributed, and scalable machine learning and predictive analytics platform. It allows users to build models on data without little expertise in model building or data sciences. It can also be used for model tuning and easy integration with production environment. It provides multiple language bindings and APIs such as R, Python*, and Java*. It supports many data formats and sources.	<ul style="list-style-type: none"> • Pattern matching • Classical statistics • Visualization
StackNet*	StackNet is a scalable predictive analytics library implemented in Java*. It is a meta-modeling methodology to generalize stacking on many levels. StackNet is empowered by H2O algorithms. It supports only CSV and SVMLight-sparse data format. There are numerous tutorial and Kaggle* examples out there to get you started.	<ul style="list-style-type: none"> • Classical statistics • Visualization • Prediction • Ensembling to improve algorithm outcomes
Caffe*	Caffe is an open source deep learning framework known for its speed, expressive architecture, modularity and extensibility. It was initially developed by UC Berkeley* Vision and Learning Center* and later optimized via community contributors at GitHub*. It is optimized for Intel architecture. It is released under BSD license. It is very helpful in image classification and segmentation and been widely used in academic exercises all over the world.	<ul style="list-style-type: none"> • Image classification • Video processing

Advanced Analytics Move-Up Guide

<u>RapidMiner*</u>	RapidMiner is a template-based visual data science platform written in Java programming language. It is distributed under the AGPL open source license and is available to download from SourceForge*. There is no need to write any code using RapidMiner, a user can perform any machine learning task using drag-and-drop options under its integrated environment from data pre-processing, data loading and transformation (ETL), data cleaning, statistical modeling, text mining, predictive analytics, visualization, evaluation, deployment and rapid prototyping.	<ul style="list-style-type: none"> • Visualization • Data cleaning • Statistical modeling • Text mining • Customer segmentation and intent detection • Customer support optimization • Risk analysis
<u>KNIME*</u>	KNIME is an open source data analytics platform with 1500 modules and GUI based features for data extraction, transformation and loading. It also provides reporting, integration, machine learning and wide data science capabilities. It is written in Java and can be extended by available plugins. KNIME has a comprehensive range of machine learning algorithms available and can connect with all major file formats and databases with native and in-database blending options. It supports an array of data types including XML, JSON, images and documents. And can integrate with other platforms.	<ul style="list-style-type: none"> • Pattern recognition • Text mining • Time-series analysis • Social media analysis/ intelligence • Classification • Association analysis • Churn prediction
<u>Weka*</u>	Weka (Waikato Environment for Knowledge Analysis) is an open source machine learning software for data mining tasks written in Java. It's an easy to use independent platform that provides flexibility for scripting experiments. It contains modules for data pre-processing, supervised learning, unsupervised learning, association rules, clustering, classification, feature selection and visualization. Weka *uses Attribute Relation File Format (ARFF) for data analysis, but also supports CSV and database using ODBC.	<ul style="list-style-type: none"> • Pattern recognition • Classification • Prediction
<u>NLTK*</u>	Natural Language Toolkit* (NLTK) is pool of libraries and programs for symbolic and statistical natural language processing. It is written in Python. It provides tools for human language data for data mining, machine learning, data scrapping, sentiment analysis and other language related tasks in the field of empirical linguistics, cognitive science, machine learning, artificial intelligence and information retrieval. It has an easy-to-use interface with over fifty corpora and lexical resources like WordNet* and other suite of text mining libraries for classification, stemming, tagging, tokenization, parsing, semantic reasoning, and understanding.	<ul style="list-style-type: none"> • Language processing • Text mining • Classification • Sentiment analysis • Text authentication • Writing analysis