opentext™

WHITE PAPER

Demystifying data science

How organizations can benefit from artificial intelligence and advanced analytics

opentext™

Contents

opentext™

Summary

By 2020, *Forbes* estimates that 85 percent of customer interactions will be managed without a human.[1] While many of us use AI technology, such as Alexa and Siri, as part of our daily lives, we may not be aware of its greater uses. In fact, with machine learning applied, AI can help teach computers, target ads and personalize content for consumers to ensure better and more informed business decisions.

This paper will clarify some key definitions around artificial intelligence and machine learning. It will also simplify some common techniques in machine learning, such as supervised learning, natural language processing and classification, and identify the types of business questions these techniques can answer.

While understanding a small number of customers may not pose a challenge, keeping pace as organizations grow and expand their customer base can be difficult. Data analytics can help reveal trends and metrics that would otherwise be lost among the masses of information. Organizations are now starting to leverage descriptive, diagnostic, predictive and prescriptive analytics to address the growing needs and demands of their customer base.

The promise of artificial intelligence is exciting but before jumping in organizations need the right data literacy, infrastructure and expertise. This paper will also cover key competencies organizations need to get started with AI and how to progress from data collection, exploration and analytics to artificial intelligence.

Finally, this paper will help define meaningful and high value use-cases with a structured framework to gather and align business, technology and data requirements for a successful artificial intelligence implementation.

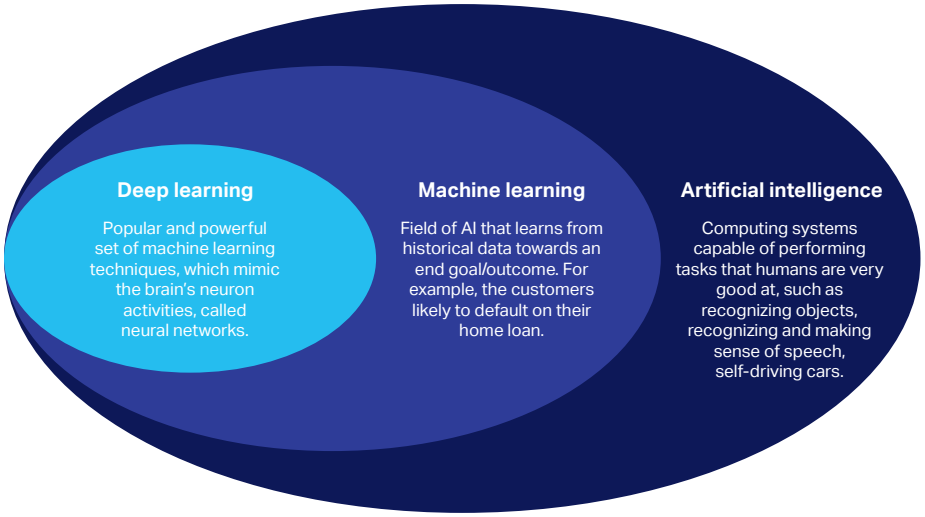
opentext™

What is artificial intelligence and machine learning?

Just as a human goes through the process of driver training to become proficient, a computer learns from experience or, more specifically, data.

According to Gartner, artificial intelligence will generate $2.9 trillion USD in business value and recover 6.2 billion hours of worker productivity by 2021.[2] To realize the high growth potential and costs savings from analytics and AI technology, we must demystify some key artificial intelligence, machine learning and analytics concepts.

Simply put, **artificial intelligence** systems automate and simplify tasks, such as recognizing objects, making sense of speech, etc. But, how does that lead to learning how to drive a car? A key concept of AI technology is the difference between learning and training. Just as a human goes through the process of driver training to become proficient, a computer learns from experience or, more specifically, data. Once the system has data on good driving practices and the "rules of the road", it becomes intelligent enough to make decisions in the real world. While there are more complexities in the learning, management and monitoring of such technology and solutions, this is the core of AI.



Source: https://www.kdnuggets.com/2018/11/an-introduction-ai.html

Machine learning, a subset of artificial intelligence, enables users to learn from historical data to achieve a desired outcome. It powers targeted ads, personalized content, song recommendations, predictive maintenance activities, virtual assistants and more.

Machine learning, a subset of artificial intelligence, enables users to learn from historical data to achieve a desired outcome. It powers targeted ads, personalized content, song recommendations, predictive maintenance activities, virtual assistants and more.

Machine learning can be broken down into two key phases, learning and predicting. In the learning phase, certain statistical techniques or algorithms are applied to historical data and/or previous business outcomes to generate a machine learning model. A model can be thought of as a set of rules or instructions, such as steps in a recipe, that one must follow to make a business decision.

For example, in order to approve a loan application, a loan officer will consider income, age, net worth and many other factors before making a final decision. Each attribute of the application is a rule or factor that the officer must evaluate to approve or reject the loan. Machine learning techniques follow a similar methodology, comparing various attributes, historical decisions and the outcome of similar applicants to estimate the credit worthiness of the new applicant.

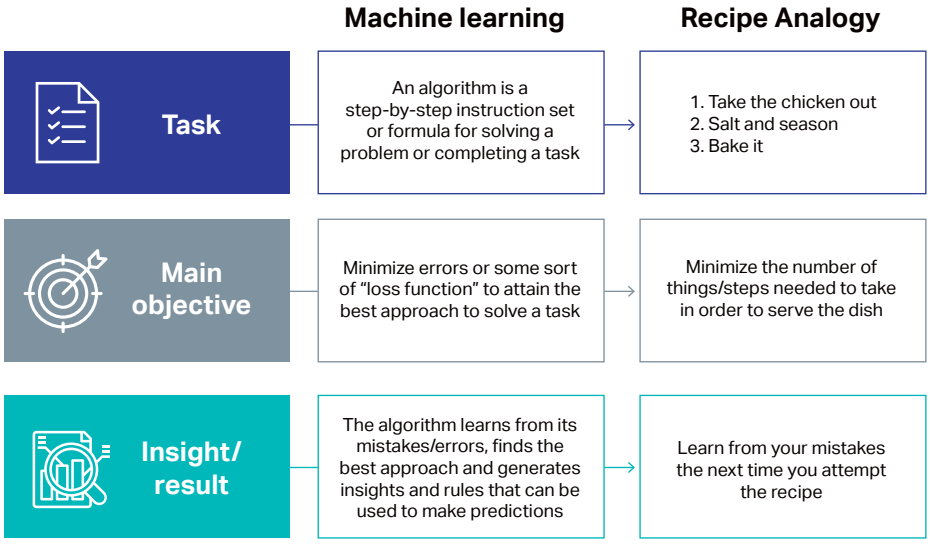
opentext™

With the growth of data, the invention of advanced algorithms and cheaper commodity hardware to process big data at scale, deep learning, a powerful set of machine learning techniques, has become prominent in the industry.

In the predicting phase, patterns identified during the learning phase are applied to new data or business processes to score or predict the likelihood of outcomes. Scoring outcomes enables organizations to optimize resource allocation and decision-making activities, make more intelligent decisions and automate key business processes at scale. Some key business questions that machine learning techniques can help answer include:

1. Will my customer purchase product X?
2. Will my customer like a recommended song?
3. Which of my customers are likely to switch to a competitor or cancel their contract?
4. Of all recently submitted claims, which ones are likely to require an additional fraud investigation unit review?
5. Is this applicant likely to default on their car loan in the future?

What do algorithms do?



With the growth of data, the invention of advanced algorithms and cheaper commodity hardware to process big data at scale, deep learning, a powerful set of machine learning techniques, has become prominent in the industry. **Deep learning** techniques mimic the brain's neuron activities, which is why they are also referred to as neural networks. Some common applications include natural language processing, image recognition, realistic photo and video generation.

opentext™

How can an organization derive business value from AI and analytics?

There are some common questions organizations consider when appealing to their customer base: Who are the customers? What do they want? How can the organization provide the best customer experience to gain a competitive advantage? Data analytics help answer these business questions.

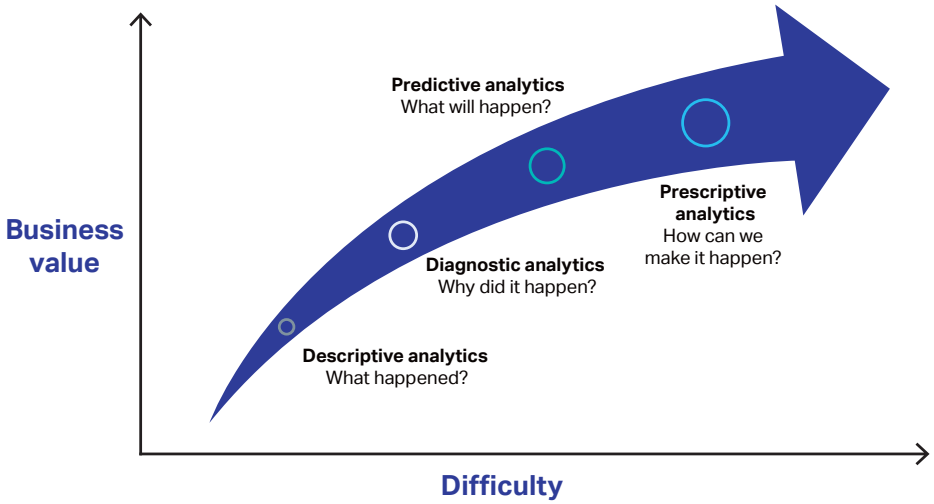
Data analytics is the science of analyzing raw data to draw conclusions from that information. Data analytics techniques can reveal trends and metrics that would otherwise be lost in a mass of information. This information can then be utilized to optimize processes to increase the overall efficiency of a business or system. Data analytics techniques can be broken down into four main types based on the difficulty of analysis and business value.

a. **Descriptive analytics** parses raw historical data and draws conclusion that help managers, investors and others determine why business changes occurred.

b. **Diagnostic analytics** provides an understanding of why events took place by examining data. A type of advanced analytics, techniques include data discovery and mining, correlation analysis and drill-down.

c. **Predictive analytics** uses statistics and modeling to predict future behavior. Using data patterns, predictive analytics identifies when patterns are likely to reoccur to identify and prevent potential risks, take advantage of future opportunities or advantageously reallocate resources.

d. **Prescriptive analytics** uses machine learning to analyze raw data to help organizations make better decision and take a proper course of action. Factoring in possible scenarios, available resources, past performance and current performance, prescriptive analytics help determine the best course of action in a situation.



opentext™

"Not going to the top is like an insight engine working at half capacity, not using all its potential."

What are the requirements for adopting AI?

This hierarchical pyramid explains the competencies every organization requires to ensure a successful AI implementation.

Data collection. At the bottom of the pyramid is data collection. At this stage, the goal is to identify what data is needed and what is available. If it is a user-facing product, are all relevant interactions logged? If it is a sensor, what data is coming through and how? Without data, no machine learning or AI solution can learn or predict outcomes.

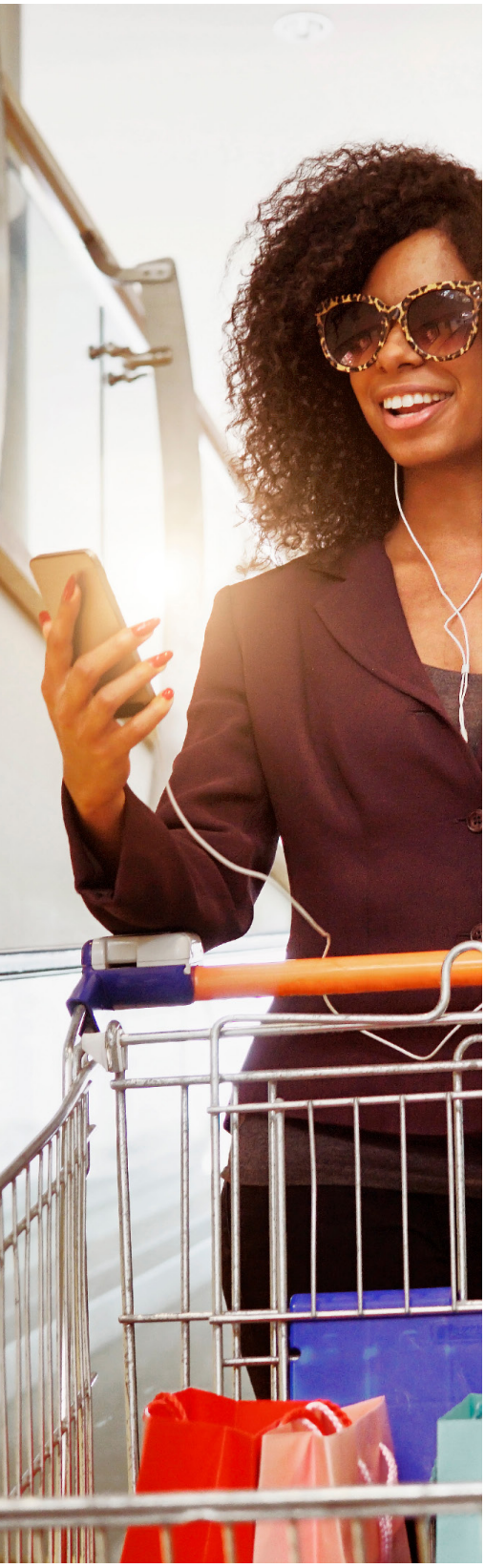
Data flow. Identify how the data flows through the system. Is there a reliable stream/ETL process established? Where is the data stored, and how easy is it to access and analyze?

Explore and transform. Only when data is accessible can it be explored and transformed for modelling. This stage is one of the most time-consuming and underestimated of the data science project lifecycle. It is at this stage that teams and organizations realize that they are missing data, their machine sensors are unreliable, they are not tracking relevant information about customers and other key issues. It forces them to return to data collection and ensure the foundation is solid before moving forward.

Business intelligence and analytics. When teams can reliably explore and clean data, organizations can start building what is traditionally thought of as business intelligence or analytics, which includes defining key metrics to track, identifying how seasonality impacts product sales and operations, segmenting users based on demographic factors, etc. However, as the goal is to build an artificial intelligence solution, it is important to start thinking about the features or attributes to include in machine learning models, what training data the machine will need to learn, what to predict and automate and how to create the labels from which the machine will learn. Label creation can be done automatically, such as when the machine breaks down and it automatically registers an event in the back-end system. Or, it can be done by introducing humans. For example, an engineer reports an issue when a machine part seems to be faulty during a routine inspection and the result is manually added to the data.

Machine learning and benchmarking. Although there is sample data that can be used to make predictions, work is not complete. A/B testing or experimentation framework needs to be in place to deploy models incrementally and avoid real world disasters. Model validation and experimentation approaches provide a rough estimate of the effects of changes before practical implementation. At this stage, a very simple baseline or benchmark for performance tracking should be established. An example fraud detection system includes monitoring high risk credit card transactions that were proved to be fraudulent and comparing them with the current operational performance of machine learning models to accurately detect fraud.

Artificial intelligence. At this stage a team might be looking to make improvements in production. This can be achieved by learning new methods and techniques in machine learning and deep learning to improve processes, predictions, outcomes and insights. By leveraging advanced and new techniques, teams can gain an Information Advantage from massive amounts of data, explore and model it faster and build solutions, such as voice assistants.



How can data science, artificial intelligence and analytics help transform business processes?

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Because there are a number of different techniques and methodologies, it is often difficult to narrow down the scope of how data science can impact business performance, operations, customer experience and costs. Here are just a few ways data science can be leveraged.

Augment employee decisions with data-driven insights and intelligence

Utilize subject matter experts' knowledge of how employees make business decisions and transform the steps into data point, applying machine learning techniques to identify the decision making pattern from this historical data to predict future business outcomes. Organizations can design an intelligent system that can handle complex requests or tasks, provide intelligent/best fit decisions for individual scenarios and empower employees to make decisions quickly and more effectively. Some example uses include credit risk scoring, automated underwriting, wealth management fund assistants and customer service chatbots.

Automate and improve the efficiency of operations with intelligent, data-driven decisions

Leverage AI and analytics techniques to drive operational efficiency. By utilizing sensor information from machines, machine learning can help predict when a specific machine is likely to require maintenance, allowing technicians to be proactive rather than reactive in maintenance efforts. Some AI applications in this context include predictive maintenance, recommender systems, robotic process automation and airline scheduling.

Apply data driven insights to make timely and consequential tactical and strategic decisions

Better inform management and strategic decisions by leveraging machine learning and advanced analytics. These tend to be ad hoc projects or solutions, where the goal is to apply statistical techniques to gain key insights around business processes. For example, by measuring analytics related to cleanliness, customer service, overall satisfaction, etc., an amusement park operations manager can determine the likelihood of repeat customers, identify key gaps in operations and better market the value of the park to the right demographics.

Personalize customer experiences

Identify and recommend personalized products at scale with recommender systems. The likes of Google, Amazon, Facebook and Apple have made personalization an expectation. Recommender systems are one of the more popular examples of how machine learning can be leveraged to analyze data across millions of users to accomplish this. By analyzing and tracking various customer touchpoints, some retailers are now able to predict the likelihood of users buying future products. It is important to note that machine learning solutions need not be 100 percent accurate to realize business value and ROI. The goal should be to conduct data-driven decision making at scale to reduce operational costs and optimize resources and targeting efforts.

opentext™

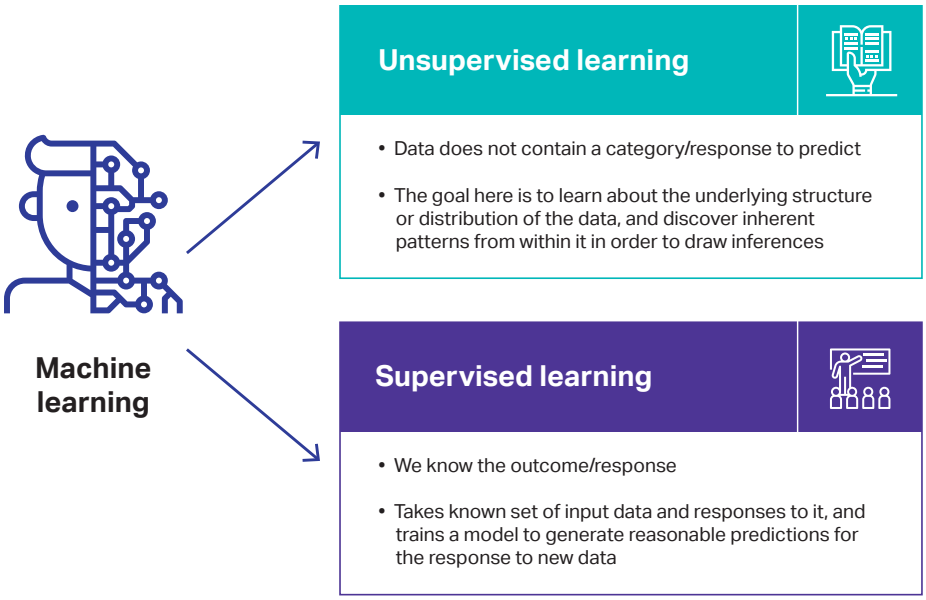
Utilize data driven insights and intelligence to accelerate new product development

Convert massive amounts of big data into meaningful and actionable insights. Voice assistants, autopilot features and smart home devices have become part of day-to-day life. This new class of AI-driven products are powered by machine learning and advanced analytics techniques, allowing organizations and teams to better understand consumer needs and wants, feature requests and usage patterns.

Common techniques and methodologies in machine learning

Machine learning takes what comes naturally to humans and applies it at scale. For example, when machine learning reviews a loan application, it can review 5,000 credit transactions, three credit reports, 10 incidents, the five-year income history of Joe Adams in seconds. This would not be possible by a domain expert. They simply do not have the capacity to reviews with the speed of a machine and provide a decision on his loan as soon as it is submitted. Even if the expert is highly experienced and efficient, it takes considerable time to review application details and there is still room for human error. Machine learning uses past experience and trends in historical data related to customers in both good standing and those that defaulted on loans to make a decision. With the combination of machine learning and good quality data, organizations can quickly make unbiased, data-driven decisions at scale in seconds.

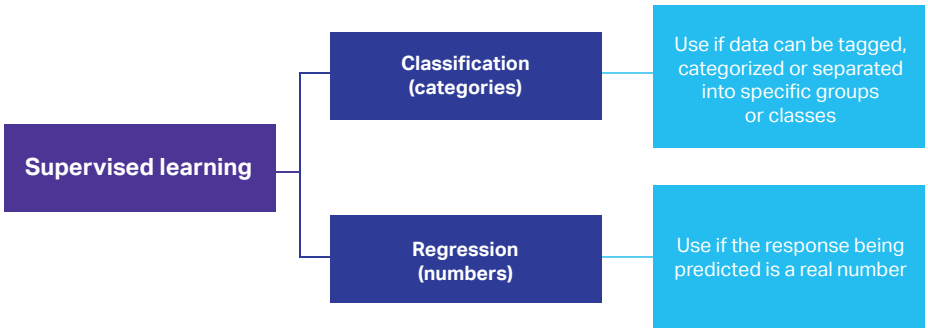
Machine learning offers various approaches to solve business problems. The first approach is based on whether there is data related to the outcome of a process. Did the machine stop working? Did the customer leave? Did the employee quit? It is important to understand and model how behavior and fluctuations in data lead to a certain business outcome. This type of machine learning is known as supervised learning.



If there is no response or category to predict, the goal is to learn the underlying structure of the data and discover patterns to draw real world inferences. For example, **unsupervised learning** approaches are commonly used to segment customers based on demographic, behavior and past product purchase history. This allows an organization to learn more about their customers, which products are frequently bought together and how different groups prefer certain services and products over others. It may not immediately understand that Emily from Philadelphia falls within customer segment X, but the organization can learn how many of its customers are similar to Emily based on behavior and consumption characteristics. Are they active on mobile? Do they use social media? Do they visit retail stores for purchases? Are they affluent? These insights can allow organizations to make data-driven decisions for future marketing campaigns, product development, etc.

Supervised learning

Supervised learning can be broken down into two categories based on what it is trying to predict.



Classification algorithms or approaches are used when asking questions regarding categories. Examples include:

- Will this customer switch to another competitor in the next month?
- Will this customer default in the next month, six months or year?
- Is an email spam or genuine?
- Is this document for compliance, legal or customer support?

Regression algorithms or approaches are used when asking questions with numerical outcomes:

- What will the temperature be at 6:00 pm today?
- In how many days will this machine stop working?
- What should be the price of a property based on size, number of rooms and location?
- How many orders am I likely to receive in the next three months for my product?

Unsupervised learning

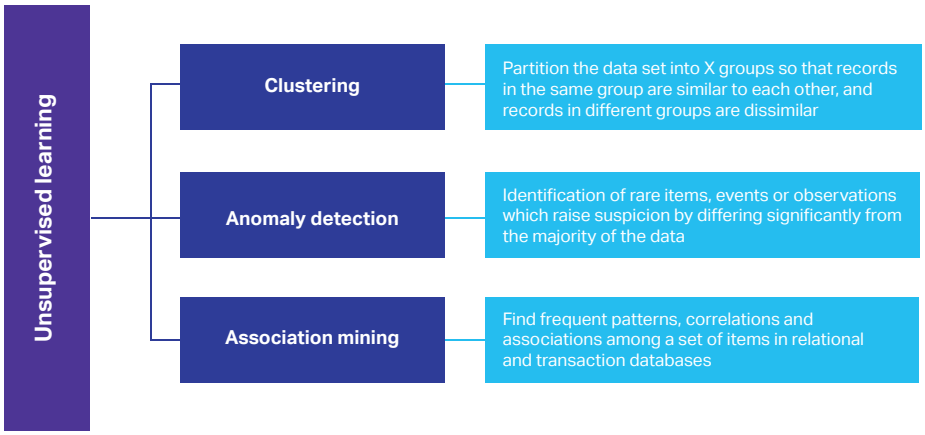
There are multiple unsupervised learning approaches and techniques that can be utilized to gain meaningful insights. One of the more popular techniques is **clustering**, which groups things that are similar or have features in common. Organizations use clustering techniques to answer business questions, such as:

- How many distinct customer groups exist for my products? Who belongs to which group?
- To which customer subgroups should I market my product and how should I target them? What are the key characteristics of each group?
- How can I group my documents into distinct categories?

If a business is looking to answer questions around the identification of anomalies or rare behavior and occurrences, **anomaly detection** techniques are utilized to identify unusual patterns that do not conform to expected behavior, called outliers. It has many applications in business, from intrusion detection, such as identifying strange patterns in network traffic, to system health monitoring, including spotting a malignant tumor in an MRI scan. Some additional questions that can be answered using these techniques include:

- Given a massive database of financial data, which transactions are suspicious and likely to be fraudulent?
- Given the huge number of container shipments arriving at a country's ports every day, which should be opened by customs to prevent smuggling, terrorism, etc.?
- Given a log of all the traffic on a computer network, which sessions represent attempted intrusions?

Association mining, another set of techniques, can help find correlations between different products or factors in an organization's data. For example, if a customer purchases baby diapers, he or she has a 60 percent chance of purchasing baby lotion within the next month. By identifying such insights using association mining, retailers can predict the need for new products and target customers with coupons and offers before the customer even realizes they are running out of baby lotion. The most common application of association mining algorithms is in market basket analysis.



Natural language processing (NLP)

Natural language processing is a set of systematic processes for intelligently and efficiently analyzing, understanding and deriving information from text data. It can organize massive amounts of text data and perform numerous automated tasks, such as automatic summarization, machine translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition and topic segmentation.

Key questions to ask and how to define high value use-cases

To identify meaningful and high value use-cases for teams and organizations, it is important to gather relevant information and requirements on three key pillars.

1. Business knowledge

- **What is the current business process?** How are things done currently? Does someone manually identify which products to recommend to each customer? Does someone manually review each loan application for fraud or risk? Does an engineer manually inspect all machinery each week for failure? Be as specific and detailed as possible in defining the current process.
- **Define measurable goals and objectives.** Is it to replace or enhance an existing process? Is it to increase revenue and conversions from product upsell and cross sell opportunities or to increase software subscriptions by three percent this quarter?
- **What areas can be improved?** What is the business question to be answered with analytics? Leverage subject matter experts to identify key pain points and gaps in the current business process. Determine what part of the process can be enhanced. Where can data-driven insights be used? Is the objective to speed up loan application processing? Identify high risk transactions on credit cards? Understand customers better? Specify key challenges and areas for improvement.

2. Solution vision

- **Why is it important to solve the current business problem/use-case?** Define what success would look like. Specifically, in order to execute a successful project, what are the minimum requirements and success criteria?
- **Define what decision or business process will be affected by the analytical solution.** Who will be affected by this tool? Who are the users of this tool? Will this impact the marketing department and analysts? Will it impact planning and maintenance personnel? Will it impact the claims processing unit of an insurance company that is responsible for mitigating fraud?
- **How is the ROI of AI and analytics measured?** Is there any current method to track/benchmark the performance of current business processes and outcomes?

3. Data adequacy

- What data is available? Is it structured or unstructured? Is there data relevant to answering the business problem? Example: Operational data is required to predict when a machine will fail.
- How much data does the organization have? Where is the data stored?
- Is data readily accessible for analysis and modelling?

opentext™

Learn more

OpenText™ Magellan™

OpenText™ Magellan™ product overview

OpenText AI white paper

OpenText™ Magellan™ infographic

Join the conversation

Keep up to date

Watch the videos

Use-case evaluation worksheet

Section 1: Business knowledge

Section 2: Solution vision

Section 3: Data adequacy

Tips and tricks

- Framing the right business question is key to success.
- Identify what success means and what the end solution will look like at the start.
- Remember that AI applications have a very different lifecycle—training, testing, modelling, experimenting and creating.
- Start small, start early!
- Iterate, iterate, iterate!

About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: opentext.com.

Connect with us:

- OpenText CEO Mark Barrenechea's blog
- Twitter | LinkedIn

[1] Forbes, *5 Ways AI Is Transforming The Customer Experience*, April 16, 2019.
[2] Gartner, *Gartner Says AI Augmentation Will Create $2.9 Trillion of Business Value in 2021*, August 5, 2019.

opentext.com/contact