

# A business guide to modern predictive analytics



IBM

# What's inside

Why this guide?	<a href="#">03</a>
The big picture	<a href="#">04</a>
Why predictive analytics and AI matter	<a href="#">05</a>
The tipping point for AI adoption	<a href="#">06</a>
How can AI augment your business?	<a href="#">08</a>
Climbing the AI ladder	<a href="#">13</a>
What are your solution options?	<a href="#">14</a>
Taking the next step	<a href="#">16</a>
Key takeaways	<a href="#">17</a>
Why combine decision optimization?	<a href="#">18</a>
Glossary	<a href="#">19</a>



## Why this guide?

In business, foresight is everything. If you can predict what will happen next, you can do the following tasks:

- Make smarter decisions
- Get to market faster
- Disrupt your competitors

Modern predictive analytics can empower your business to augment historical data with real-time insights then harness this to predict and shape your future.

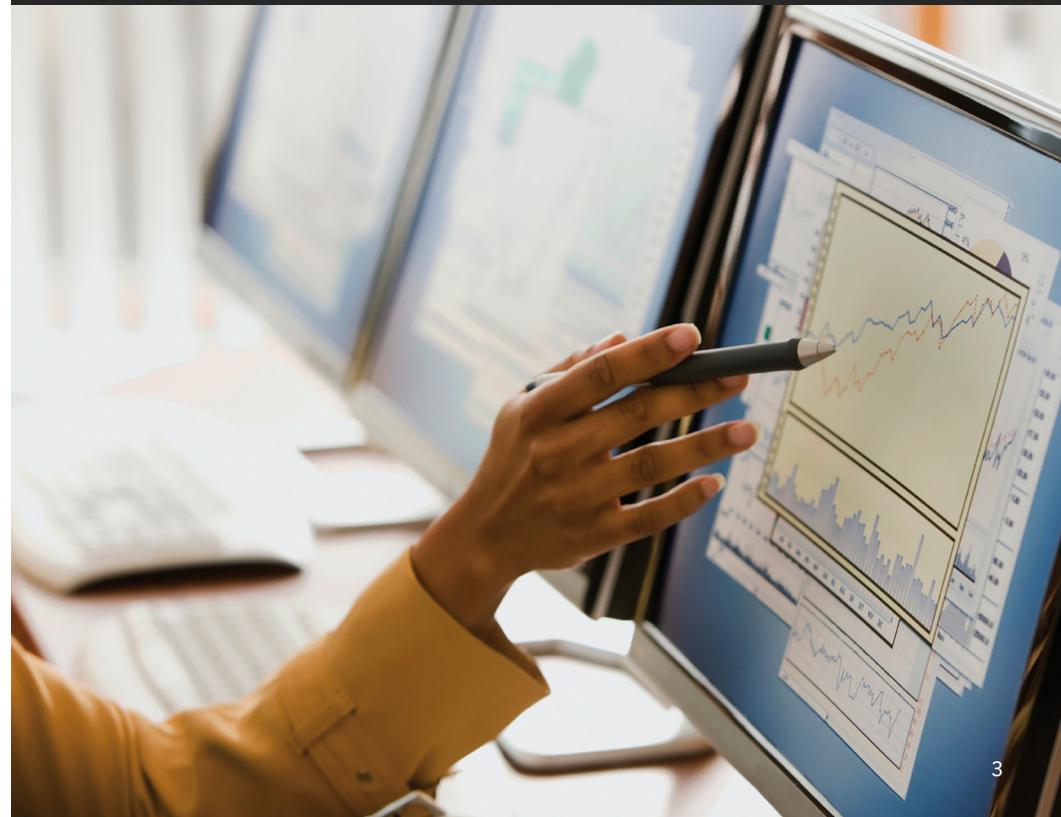
Predictive analytics is a key milestone on the analytics journey—a point of confluence, where classical statistical analysis techniques meet the new world of artificial intelligence (AI).

According to [Forrester Research](#), enterprises have reached a point to begin combining machine learning with knowledge engineering. Augmenting data with human wisdom will dramatically accelerate the development of AI applications.

This guide will help your business perform the following actions:

- Navigate the modern predictive analytics landscape
- Identify opportunities to grow and enhance your use of AI
- Empower both data science teams and business stakeholders to deliver value, fast

Modern predictive analytics is about using machine-generated predictions with human insight to drive business forward.



# The big picture

As the AI revolution takes hold, businesses are increasingly asking their data science teams to tackle the big questions.

As a result, data scientists are expected to do much more than work on one-off research projects. They need to find repeatable, automated ways to provide real-time insights for day-to-day decision-making.

To meet these expectations, data science leaders not only need to be able to explain the potential of modern predictive analytics technologies to business stakeholders—they also need to deliver the results.

The ability to define and execute a successful data science strategy will be one of the key differentiators between leaders and followers in the years ahead.

This is no simple task. Building up your data science capabilities will involve the following activities:

- Attracting and retaining a disparate team of skilled specialists
- Empowering them to collaborate seamlessly
- Putting sound governance structures in place to ensure that predictive models can always be trusted by the business

Above all, data science and business teams need to find new ways to collaborate effectively. These methods include understanding what predictive analytics can do and identifying the areas where AI will drive business advantage.

- 1 How do our customers behave?
- 2 Why are our markets fluctuating?
- 3 What makes our business strategies succeed or fail?
- 4 What will happen next?
- 5 How are the projected funded?
- 6 Where are the buying centers?



## Why predictive analytics and AI matter

Predictive analytics is not a new concept. Statisticians have been using decision trees and linear and logistic regression for years to help businesses correlate and classify their data and make predictions.

What's new is that the scope of predictive analytics has broadened. Breakthroughs in machine learning and deep learning have opened up opportunities to use predictive models in areas that have been impractical for most business investments—until now.

Enterprises are seeing an unprecedented confluence of intuitive tools, new predictive techniques and hybrid cloud deployment models that are making predictive analytics more accessible than before.

This situation has created a tipping point. For the first time, organizations of all sizes can do the following activities:

- Embed predictive analytics into their business processes
- Harness AI at scale
- Extract value from previously unexplored “dark data”—including everything from raw text to geolocational information

If you can evolve from departmental, small-group AI projects and advance toward an enterprise data science platform, your organization stands to gain significant competitive advantage. Those who don't seize the opportunity risk falling behind the curve.

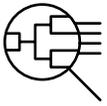
# \$77.6 billion

will be spent on cognitive and AI systems by 2022

(Source: [IDC](#))



# The tipping point for AI adoption



## What types of data can be analyzed?

### Before:

Primarily relational data at scale; other types of data require ad hoc research projects.

### Now:

Relational data, semi-structured documents, text, sensor data and more; both historical and real-time analytics are possible at scale.



## What tooling is available?

### Before:

Disparate, incompatible tools that require multiple handovers between teams with different expertise.

### Now:

A blend of drag-and-drop interfaces and open source notebooks that make collaboration between teams more convenient.



## What analytical techniques can be used?

### Before:

Basic statistical techniques such as logistic and linear regression.

### Now:

Statistical techniques augmented with state-of-the-art machine learning and deep learning algorithms.



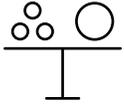
## How do enterprises deploy analytics applications?

### Before:

Applications and analytics are tied to data on-premises servers and data warehouse appliances, reducing opportunities for anytime, anywhere analytics.

### Now:

Hybrid, multi-cloud deployments help push analytics to wherever data resides, while combining on-premises security with flexibility and scalability.



## How can enterprises integrate analytics into our business processes?

### Before:

Generate static reports for manual analysis by business experts.

### Now:

Seamlessly embed predictive models into new apps and enterprise applications.



## How do enterprises implement governance?

### Before:

Ad hoc adherence to policies at departmental level, with minimal visibility or traceability.

### Now:

A coherent governance and security framework enables enterprise-wide policies to be enforced at scale.



## How do enterprises inject artificial intelligence into modern applications?

### Before:

A total disconnect between application development and data science teams means each deployment is a custom process.

### Now:

The data science lifecycle is designed to create a standardized, repeatable process for AI integration.



## How can enterprises progress on their analytics journey?

### Before:

Each step from descriptive to predictive and prescriptive analytics requires separate tools, skills and investment.

### Now:

An integrated platform supports analytic progression, simplifies onboarding and grows with you as your needs change and skills develop.

## How can AI augment your business?

In theory, adopting a modern approach to predictive analytics should be straightforward. The technology is no longer an obstacle, and better tooling is lowering the barriers to entry significantly.

However, in practice, delivering value can still be a challenge. It's especially easy for business stakeholders to get caught up in the hype around AI and have unrealistic expectations of what data science can achieve.

### Defining use cases

The first task for data science and business leaders is to work together to identify concrete, practical use cases where modern predictive analytics can deliver value.

Some use cases may be generally applicable across most industries, such as the following examples:

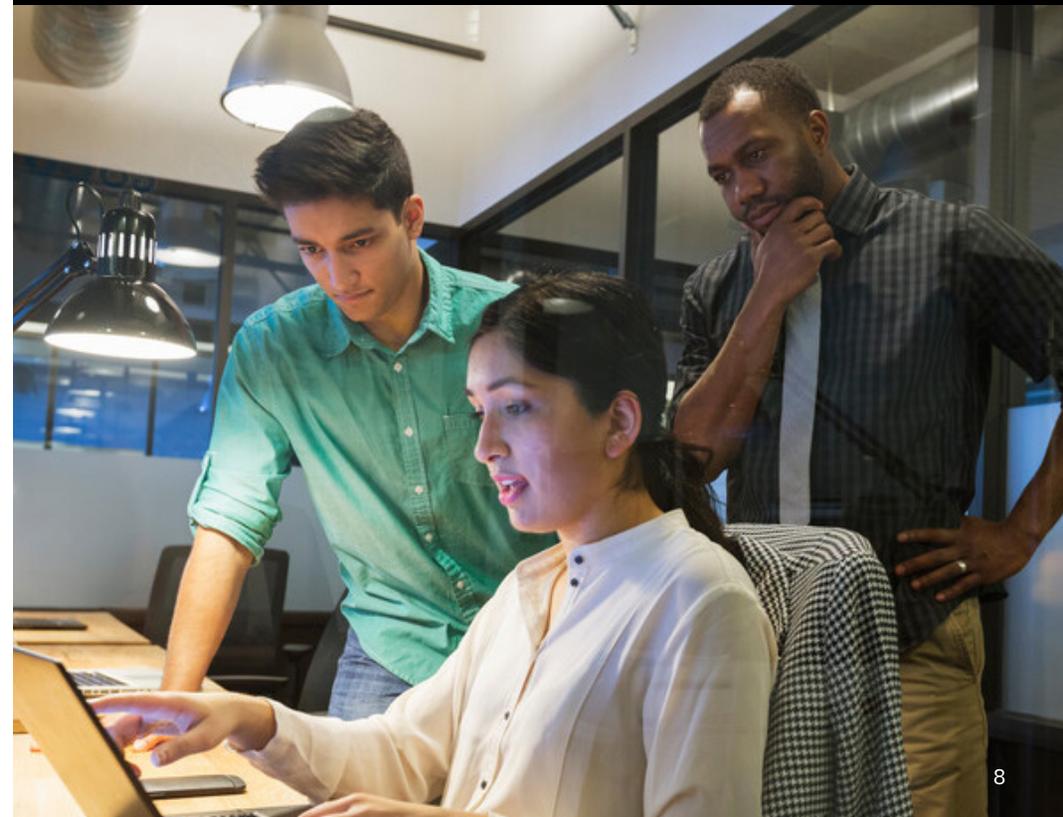
- Product recommendation and “next best action” models for sales and marketing teams
- Contact center automation for customer support teams

Other use cases may be specific to a particular industry, department or even team within a business. These tend to be more difficult to execute, but they have a greater potential to unlock unique competitive advantages.

Which business functions are leading business investment in AI systems?

**46%** sales and marketing  
**40%** customer support

(Source: [Forrester Research](#))



## General use cases

When a business begins investing in a new technology, it often makes sense to pick the lowest-hanging fruit first.

Predictive analytics is no different. Several use cases are widely applicable across industries, and vendors have already developed general-purpose, prepackaged models and services.

These services can be an excellent starting point for businesses that want to transform data science from a research function into an embedded part of day-to-day operations. They are easy to deploy, require minimal custom development and deliver value quickly.

## Contact center optimization

Handling unpredictable volumes of customer calls, emails, SMS and chat messages is a challenge for many customer service teams.

Intelligent chatbots are a powerful and cost-effective way to take the pressure off employees and reduce wait times for customers. These chatbots use the following features to understand customer inquiries:

- AI-powered speech recognition
- Natural language processing
- Content analytics to explore the company's knowledge base and find helpful answers, without needing human intervention

## Some of the most common cross-industry use cases for what modern predictive analytics can provide include:

- **Increasing cross- and up-selling** with personalized real-time recommendations and offers
- **Boosting loyalty** by anticipating customer churn and intervening to prevent it
- **Optimizing offerings** by listening to voices of customers and anticipating future needs
- **Enhancing marketing** with targeted, personalized campaigns
- **Minimizing inventory costs** and improving resource management with accurate forecasting
- **Improving productivity** by allocating the right employees to the right jobs at the right time and creating accurate labor forecasts
- **Reducing maintenance costs** by anticipating faults before they occur
- **Mitigating risk** with accurate customer credit scoring
- **Detecting fraud** by identifying suspicious behavior patterns
- **Unlocking new business models** by addressing untapped demands and integrating prediction into modern apps



## Industry-specific use cases

Innovative organizations across many industries are already investing in building their own predictive models to solve specific business problems. The next two pages highlight just a few of the potential applications for AI and predictive analytics across several major industries.



## Commercial banking

Commercial banks use predictive analytics for the following tasks:

- Assess market and counterparty risk on trades
- Assess credit risk for loan applications
- Detect fraudulent transactions in real time
- Harness predictive modeling to accelerate loan approval processes



## Insurance

Insurers use predictive analytics for the following tasks:

- Detect fraudulent claims
- Optimize quotes and premiums by assessing relevant risks for each applicant
- Predict hazardous weather events to reduce auto insurance claims



## Energy and utilities

Utilities use predictive analytics for the following tasks:

- Manage vast networks of physical assets
- Forecast production and demand patterns
- Predict outages before they happen
- Plan for supply and demand



## Government

Governments rely on accurate statistics to inform policy-making across many areas, including the following use cases for predictive analytics:

- Detect benefit fraud
- Predict usage patterns for public services
- Optimize waste management and traffic flows



## Manufacturing

Manufacturers use predictive analytics for the following tasks:

- Keep production lines running smoothly by modeling product quality and detecting defects
- Optimize warehouse management and logistics
- Develop sensors for autonomous vehicles by using machine learning models



## Retail

Retailers use predictive analytics for the following tasks:

- Manage customer loyalty programs
- Boost cross- and up-selling by making targeted recommendations based on customer profiles and sophisticated propensity models
- Enable accurate demand forecasting



## Food

The food industry uses predictive analytics for the following tasks:

- Automate data collection and analysis on food health
- Predict and warn of potential health outbreaks to enable rapid intervention
- Protect companies' sensitive data, making it safe for competitors to collaborate



## Healthcare

Healthcare organizations can use statistical modeling techniques for the following tasks:

- Monitor streams of data from ECGs and other medical devices
- Predict when a patient's condition may change
- Perform medical research
- Analyze streams of patient data in real time



## Retail banking

Retail banks use predictive analytics for the following tasks:

- Enhance customer satisfaction through faster credit scoring
- Combine flexibility with robustness and security through hybrid cloud infrastructure
- Cut costs and accelerate development due to innovative architecture



## Transportation

Transportation and logistics companies use predictive analytics for the following tasks:

- Optimize route planning
- Enable predictive maintenance for vehicles
- Optimize supply chain operations

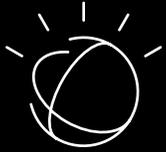


## Education

Education institutions use predictive analytics for the following tasks:

- Predict student achievement and retention
- Identify students who need extra support to reach their goals
- Strengthen donor relationships
- Track student movements to help reduce absenteeism

# Climbing the AI ladder



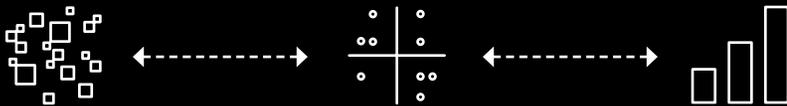
**Infuse** - Operationalize AI with trust and transparency

**Analyze** - Scale insights with AI everywhere

**Organize** - Create a trusted analytics foundation

**Collect** - Make data simple and accessible

**Data of every type, no matter where it lives**



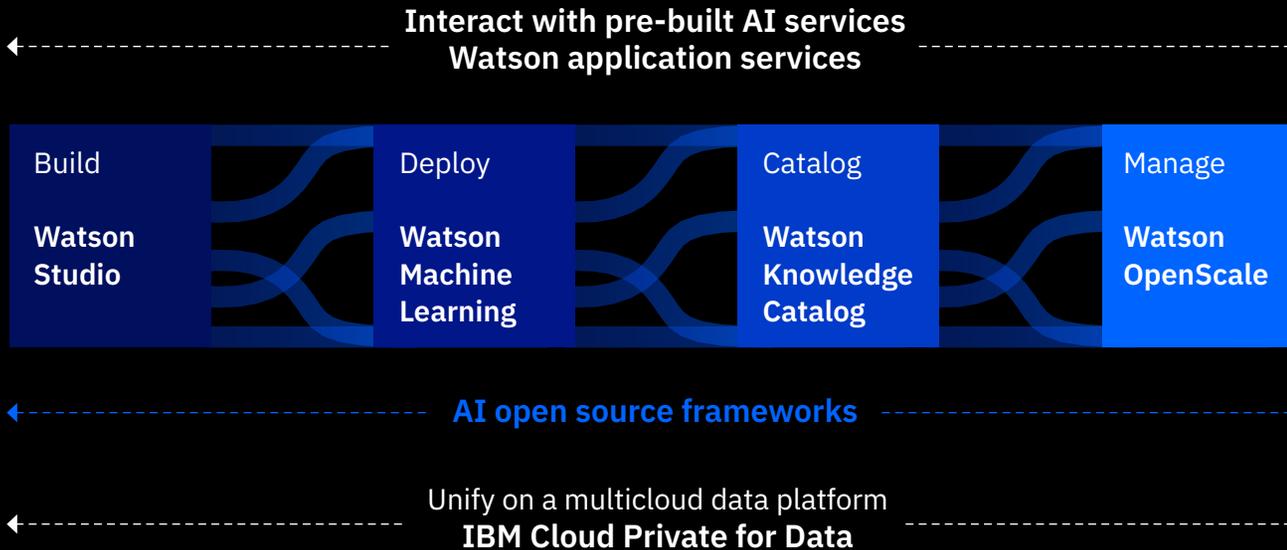
Achieving success with modern predictive analytics is a journey. It's important to pitch AI strategy at the right level for a business, taking both technical and organizational maturity into account. Data science and business leaders need to work together to define the best and fastest way to deliver business value.

From the technical perspective, you can visualize AI maturity as a ladder. The first step on the ladder is data collection, because without data, you won't have anything to analyze or model. The next step is data organization. Add metadata for governance and discoverability, to ensure that the right data is always available to the data scientists who need it.

While data collection and organization are important topics, they're beyond the scope of this guide. Instead, let's focus on helping climb the following top two levels of the ladder:

- Analyzing data by building, training and testing predictive models
- Infusing AI into operations by deploying those models into production as part of your applications

# What are your solution options?



The AI portfolio from IBM offers everything you need to reach the top rungs of the AI ladder.

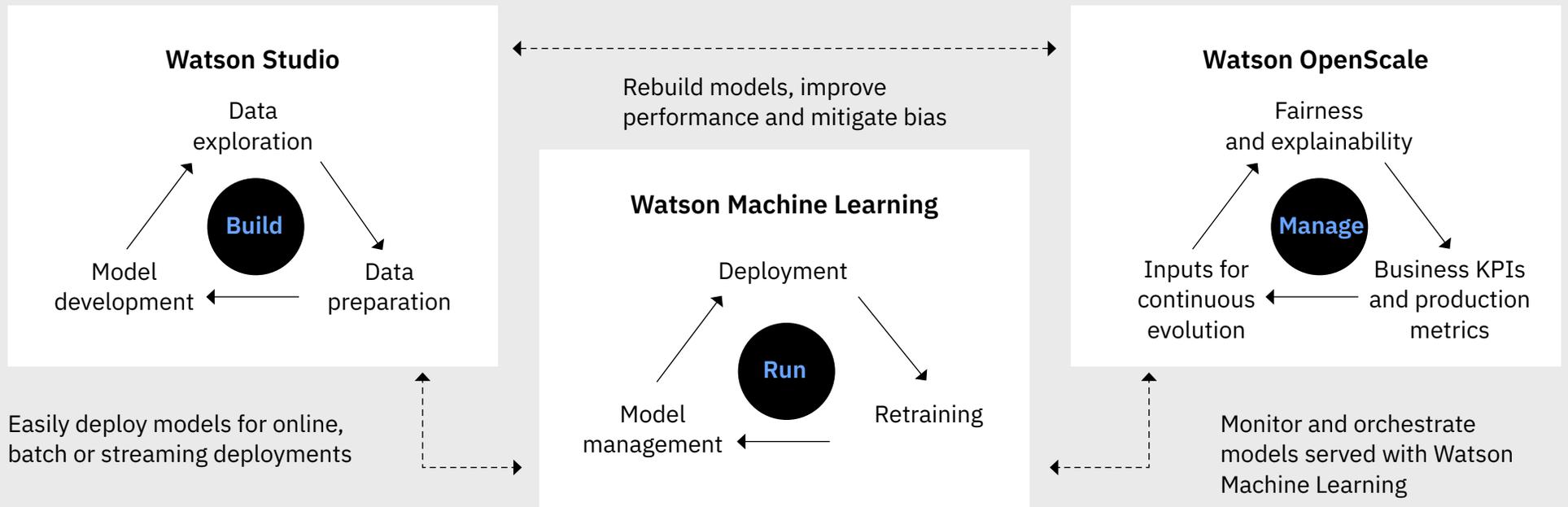
Pre-built AI services such as [Watson Assistant](#) and [Watson Visual Recognition](#) help you address common use cases quickly and efficiently, delivering value fast.

When you're ready to start developing your own AI solutions, [Watson Studio](#) and [Watson Machine Learning](#) provide seamless workflows for building, training and deploying predictive models. These solutions empower you by harnessing both state-of-the-art IBM tools and the best open source AI frameworks.

[Watson Knowledge Catalog](#) provides robust data governance and discoverability for models and data, while [Watson OpenScale](#) helps you monitor and manage models in real time—boosting accuracy, increasing explainability and mitigating bias.

[IBM Cloud™ Private for Data](#) unifies access to all these capabilities and provides a powerful multicloud data platform.

[IBM Data Science Premium add-on](#) for IBM Cloud Private for Data provides the additional data science productivity capabilities such as SPSS Modeler and Decision Optimization to accelerate the time to value and increase the chance of your AI/ML project success.



The top two steps of the AI ladder are Analyze and Infuse. To reach these steps, organizations must help data scientists and business stakeholders work together effectively at every stage of the data science lifecycle.

The complete lifecycle can be visualized as the following three sub-cycles that interact with each other:

### Build

Data scientists explore business data to identify interesting features, then prepare well-structured data sets that are used to design predictive models.

### Run

Operations teams train, test, deploy and manage the models, and retrain them when necessary.

### Manage

Business experts monitor the models' runtime performance, look for any signs of bias or need of explanation provide feedback and notify the data science team when they need retraining.

## Taking the next step

Depending on their level of progress on the AI ladder, businesses may have different requirements based upon the level of predictive analytics adoption across their organization.

### Starting out

When businesses begin building their data science capabilities, they often start with ad hoc projects—developing models to answer specific questions or support research projects. With solutions such as Watson Studio Desktop, data scientists can work 24x7 on their own computers or laptops and sync up with a wider team when needed.

### Growing up

When data science is adopted widely, different departments need to deploy their models, connect them to data sources and infuse them into production applications. Watson Studio and Watson Machine Learning make it easier for departmental data science and IT teams to collaborate across this lifecycle.

### Going enterprise-scale

Once AI is embedded into business-critical processes, building a central platform is vital in order to manage and govern models and data. IBM Cloud Private for Data can provide the infrastructure and tools required for a comprehensive, multicloud platform that acts as a single point of control.

## Getting practical

Whether you're a data scientist or a business leader, the best way to learn how the modern predictive analytics portfolio from IBM can transform your business is to experience it for yourself. Try one of the following tutorials to get started:

### Perform a machine learning exercise

Dive into machine learning by performing an exercise in IBM Watson Studio using Apache SystemML. [Learn more](#)

### Create a scoring model to predict heart rate failure

Use IBM Watson Studio to build a predictive model with IBM Watson Machine Learning. [Learn more](#)

### Predict equipment failure using IoT sensor data

See how IBM Watson Studio can analyze multivariate Internet of Things (IoT) sensor data and predict equipment failure. [Learn more](#)

### Analyze open medical datasets to gain insights

Use IBM Watson Studio to run machine learning classifiers and compare the outputs with evaluating measures. [Learn more](#)

### Shape and refine raw data

Work with IBM Data Refinery to prepare large data sets for predictive analysis. [Learn more](#)

# Key takeaways

The modern predictive analytics portfolio from IBM offers the following benefits data science and business leaders can use to help seize competitive advantage in the age of AI:

## Scale

- Reduce operational workload and costs by automating data science and data engineering tasks
- Train, test and deploy models seamlessly across multiple enterprise applications
- Extend common data science capabilities across hybrid, multicloud environments

## Speed

- Accelerate development by harnessing pre-built applications and pre-trained models
- Deliver value faster by helping data science and business teams collaborate
- Streamline model building by combining state-of-the-art IBM and open source software

## Simplicity

- Take advantage of a central platform to manage the entire data science lifecycle
- Standardize development and deployment processes
- Create a single framework data governance and security across the organization

Watson Studio helps businesses focus on solving problems and identifying opportunities.

[Learn more](#)

Watson Machine Learning empowers businesses to deploy and manage models to give the results they need fast.

[Learn more](#)

# Why combine decision optimization with predictive analytics?

IBM Decision Optimization is a prescriptive analytics solution that enables highly data-intensive industries to make better decisions and achieve business goals by solving complex optimization problems. Business leaders use this tool to improve their efficient use of resources, including but not limited to the following activities:

- Inventory flow for supply chain
- Workforce scheduling
- Routing of transportation

This solution works well with predictive analytics by using the predictive outcomes of machine learning applications to provide optimized outcomes. Machine learning provides insights on the future based on observations given by users. With machine learning, you know the answer, and you train the machine how to find that answer.

Decision optimization lets you take the next step and act on that information. With decision optimization, while you don't know the answer, you do know a lot about what is a good and bad answer. You take the output from your machine learning and specify an action for decision optimization to make, which can include optimization rules and constraints to achieve business goals.

Following that action, decision optimization returns answers to deliver value to the business, such as actionable items and recommendations for change. By performing this activity, decision optimization enhances what predictive analytics can offer you.

The solution lets teams combine optimization and machine-learning techniques with model management, deployment and other data science capabilities to develop optimal solutions that improve operational efficiency.

[Learn more](#)



# Glossary

**Algorithms** are sets of rules that define a sequence of operations that can be applied to data to solve a particular problem.

In a data science context, the term encompasses a huge range of techniques, including the following:

- Decision trees and regression models
- Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) and exponential smoothing
- Transfer functions with predictors and outlier detection
- Ensemble and hierarchical models
- Vector machine and temporal causal modeling
- Time series and spatial AR for spatiotemporal prediction
- Generative adversarial networks (GANs) and reinforcement

Your data science platform should give you easy access to all these powerful algorithms.

**Artificial intelligence (AI)** is the ability of computer systems to interpret and learn from data. The term is most commonly used to describe systems built using machine learning or deep learning models. AI techniques can be used to enable computers to solve a wide range of problems that were previously considered intractable.

**Bias** is a common issue when designing, training and testing models that can lead to inaccurate predictions. Mitigating bias by monitoring and auditing models during runtime is an increasingly important topic as businesses seek to adopt AI more widely.

**Classification models** aim to put data points into categories by comparing them with a set of data points that have already been categorized. The result is a discrete value, meaning one of a limited list of options, rather than a score. For example, a classification model can give a yes or no answer on whether customers are likely to make a purchase or if they are a bad credit risk. Classification models can be built using various techniques, including decision trees and logistic regression.

**Content analytics** is the analysis of unstructured data in documents of various formats, including text, images, audio and video files. Machine learning techniques can greatly accelerate analyzing large repositories of content that would previously have required workers hundreds or thousands of hours to review and classify.

**Data science** is a wide-ranging discipline that unifies aspects of statistics, data analysis and machine learning to harness data to solve business problems.

**Deep learning** is a branch of machine learning that uses neural networks with large numbers of hidden layers. These highly sophisticated networks are used in cutting-edge fields of deep learning such as computer vision, machine translation and speech recognition.

Training a deep neural network is extremely computationally intensive, typically requiring clusters of machines with high-performance processors. A hybrid cloud platform such as IBM Watson Studio or IBM Cloud Private for Data can make this kind of infrastructure more accessible and affordable for companies of all sizes.

**Deployment** is the process of integrating a model into your business applications and running that model against real-world data. Making and moving the model through test, staging and production environments requires collaboration between your data science, application developers and IT operations teams.

It can be challenging to integrate open source data science tools with the organization's existing continuous integration and deployment pipeline. To avoid manual deployments with multiple handovers between teams, a coherent data science platform with automated deployment capabilities can be a major advantage.

**Development** of predictive models involves the use of traditional statistical techniques or machine learning algorithms to create and refine models by training and testing them against your data sets.

The development process is highly iterative; you may need to train dozens or even hundreds of models to achieve the level of accuracy you require. That's why automating the workflows around model development and training can deliver huge value.

**Explainability** is an important attribute of any system that uses predictive models to make recommendations and assist business decision-making. A predictive model seen as complex and mysterious can be difficult to convince business stakeholders, regulators and customers to trust its output. The advanced runtime monitoring and logging capabilities of Watson OpenScale provide context around each decision, making AI models transparent and auditable.

**Exploration** of data is an important part of the model building process. This activity aims to reveal interesting features in a given data set, uncover hidden relationships and highlight use cases where predictive modeling could deliver business value.

During the exploration phase, it's critical to exercise data science skills and business knowledge to define questions you want to answer and outcomes you want to predict. This may result in an iterative cycle of preparation and exploration until you have fully explored the domain and have the data in the right shape to proceed.

**Geospatial analytics** is the analysis of geographic data such as latitude and longitude, postal codes and addresses. This analysis is extremely useful for solving many kinds of practical data science problems. A modern data science platform should make it easy to detect, parse and calculate geospatial information, and offer easy integration with mapping tools to visualize the results.

**Inference** in artificial intelligence applies logical rules to the knowledge base to draw conclusions in the presence of uncertainty. With inference, users get a prediction that is simplified, compressed and optimized for runtime performance.

**Linear regression** is a statistical process using one independent variable to explain or predict a value or score. Examples include the number of SKUs of a product sold in a given week or the percentage risk of a customer closing their account.

**Logistic regression** is a statistical process used in predicting outcomes. The process differs from linear regression in that the one independent variable has only a limited number of possible values rather than infinite possibilities. Users employ logistic regression when the response falls into categories such as numeric orders like first, second, third and so on.

**Machine learning** uses statistical techniques to derive sophisticated predictive models and algorithms from large data sets, without requiring explicit programming.

Typically, you start this iterative process by dividing a data set into two subsets for training and testing. You train your models against the training set and test their performance against the testing set with dozens or hundreds of variations to assess their predictions' accuracy. By running this process and basing the next generation of variations on the best performers from each iteration, the model gradually learns and improves performance.

The main approaches to machine learning can be divided into two categories: supervised and unsupervised learning.

**Management** of models is vital to ensure that they remain accurate over time. Retraining models regularly to take new data into account is critical, so model development, implementation, deployment and management should form a continuous cycle.

This management can be difficult to achieve with disparate open source tools. Using an end-to-end data science platform can avoid gaps in the process. The platform also can ensure that appropriate teams will be notified immediately and can take rapid action whenever a model's performance begins to degrade.

**Natural language processing (NLP)** is a field of AI that focuses primarily on enabling computers to analyze unstructured textual data. Common use cases include speech recognition, natural language understanding and sentiment analysis.

**Neural networks** provide a framework for training models that enables complex interaction between many machine learning algorithms to help identify optimal models.

The structure of interconnecting neurons in the brains of humans and other animals inspired the structure of artificial neural networks. Layers connect the artificial neurons. Data traverses the structure from the input layer through one or more hidden layers to the output layer. During this traversal, mathematical functions transform the data into a prediction whose accuracy you can assess.

**Open source software** has become an increasingly dominant paradigm in many areas of statistical modeling and machine learning. Languages like R, Python and Scala, big data architectures such as Apache Hadoop and Spark, and machine learning frameworks like TensorFlow and Spark MLlib, are all major players in the world of predictive analytics and data science.

Open source frameworks often focus on developing high-quality tools that target specific parts of the data science process, such as model development or training. As a result, they often leave the end user responsible for integrating all the tools together into a coherent workflow. This task can be a problem when you are trying to scale predictive analytics across the enterprise and embed AI into business processes.

**Predictive analytics** uses historical data to model a specific domain or problem and isolate the key factors that have driven specific outcomes in the past. Models built using this process predict likely future outcomes from new data.

Predictive analytics can encompass a wide range of techniques, from classical statistical modeling to machine learning algorithms.

**Predictive models** are algorithms that map an input, meaning a piece of data, such as a database record, text sample or image, to an output or prediction. Outputs are typically either continuous variables, such as a number or percentage, or discrete categories, such as "yes" or "no." There are two major types of predictive models: regression models and classification models.

**Preparation** of data is one of the first steps in the data science process. Most projects start by refining data sets to ensure that the quality is high enough to bear the weight of detailed analysis.

In many cases, your source data may need to be cleaned and transformed into a format that is more amenable to model and analyze. If you're building a machine learning model, you may also need to invest in manually labeling the data for use in supervised learning.

**Regression models** are useful when you have a data set that contains multiple variables and want to analyze the relationship between them. Specifically, regression models can reveal how one specific variable is likely to change when other variables are altered.

Linear regression can be used to predict a value or score. Examples include the number of SKUs of a product sold in a given week or the percentage risk of a customer closing their account. week and the percentage risk of a customer closing their account.

**Statistical modeling** is a domain of mathematics that involves the creation of models based on probabilistic assumptions about a set of data. Businesses have used statistical models to analyze important features of their data sets and identify correlations that can be used to classify data or generate predictions.

**Supervised learning** is a method of training a machine learning model using a data set where the data has already been correctly labeled. The model produces an output variable—typically a category or a value—so its accuracy can easily be assessed by comparing the output to the labeled input. Linear regression, random forests and support vector machines are all popular examples of supervised learning algorithms, and most predictive models are built using these techniques.

**Testing predictive models** is essential for determining the accuracy of data in AI processes along with training. Predictive models need to be tested continuously to improve accuracy. If a model fails, analysts must identify the root cause and retrain and test to improve the models.

**Text analytics** measures unstructured content using linguistic rules, natural language processing and machine learning. This process reviews data the same way as performed by human brains, but at a faster rate. With text analytics, you obtain more insights and discoveries from unstructured content, which makes up approximately 90 percent of all data.

**Training predictive models** is a key element of machine learning, deep learning and other AI processes to determine which data is useful. A model trained to give accurate predictions can be used to score real-time data. Models must be retrained periodically to adjust for changing behavior patterns.

**Unsupervised learning** is a method of training machine learning models with unlabeled data. The aim is typically to model and highlight interesting patterns or structures within the data. Clustering and association problems are common domains for unstructured learning—for example, finding interesting new ways to segment customers or identify similarities between them.

**Visualization** is the process of representing data graphically, often using charts and diagrams. To understand data, humans need to be able to visualize it. This process is important both when presenting your results to business stakeholders and when exploring a new data set during the early stages of a project.

Your predictive analytics platform should provide an intuitive graphical interface with visualization tools. These features help you to start making sense of even the largest data sets in minutes.



© Copyright IBM Corporation 2019

IBM Corporation  
New Orchard Road  
Armonk, NY 10504

Produced in the United States of America  
March 2019

IBM, the IBM logo, ibm.com, IBM Cloud, IBM SPSS Modeler and IBM Watson are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM DOES NOT WARRANT THAT ANY SYSTEMS, PRODUCTS OR SERVICES ARE IMMUNE FROM, OR WILL MAKE YOUR ENTERPRISE IMMUNE FROM, THE MALICIOUS OR ILLEGAL CONDUCT OF ANY PARTY.