

Predictive Analytics with MATLAB

Unlocking the Value in Engineering and Business Data

Executive Summary

Predictive analytics is the engine of evidence-based decision making. Today there are many opportunities that big data and engineering techniques are bringing to the world of analytics.

What can you do with engineering-driven analytics?

- You can predict the quality state of a polymer production machine. *Predictive Maintenance* helps improve quality and reduce downtime.
- Using sensor fusion and object detection, you can automatically stop a 40-ton truck that is traveling at 50 km/hour—in turn, avoiding accidents and saving lives.
- With sensor analytics, medical devices can predict seizures to enable proactive care and better health outcomes.

All of these examples use machine learning in novel ways to solve long-standing problems. However, successfully applying machine learning techniques requires data science skills—a special skillset with a low supply and very high demand.

Addressing the data science skills gap, McKinsey and Company predicts that within a couple years, the U.S. alone could face a shortage of up to 190,000 people with deep analytical skills.¹

And then there's the matter of deployment. With a combination of business and engineering data, there's an opportunity to deploy predictive analytics widely in traditional IT systems, as well as on embedded devices where analytics work in real time for decision support or decision automation.

This paper shows how *MATLAB*[®] users can perform predictive analytics with engineering, scientific, and field data as well as business and transactional data, deploy to large-scale production systems and embedded systems, and effectively unlock the value of data to make informed decisions.

What Is Predictive Analytics?

Predictive analytics is the process of using data analytics to make better business decisions and predictions based on data. It uses data along with analysis, statistics, and *machine learning* techniques to create a predictive model for forecasting future events.

Predictive analytics starts with a business goal. The process harnesses heterogeneous, often massive, data sets into models that can generate clear, actionable outcomes to support business decisions.

The term “predictive analytics” does not describe a particular statistical or machine learning technique, but, rather, the application of a technique to create a quantitative prediction about the future.

¹McKinsey & Company, *Big data: The next frontier for innovation, competition, and productivity*

Frequently, supervised machine learning techniques are used to predict a future value (How long can this machine run before requiring maintenance?) or to estimate a probability (How likely is this customer to default on a loan?).

Decision Support or Decision Automation

Predictive analytics involves applying robust, statistically motivated methods on complex data to understand what has happened and why, to predict what will happen in the future, and to suggest responses to achieve desired outcomes.

When the result of the analytic is a suggestion for a human, this is often called decision support. A person considers the results and decides to take action.

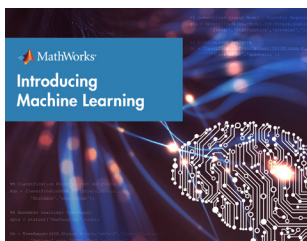
But especially for engineering, we are often interested in an automated response—we want the machine or system to take the right action itself. When the result is an action taken automatically based on the results of the analytic, this is called decision automation. For example, when a vehicle automatically applies the brakes to prevent a collision, this is decision automation. There are no people involved in the decision.

Using Big Data to Answer Big Questions

Predictive analytics is often discussed in the context of *big data* as businesses apply algorithms to derive insights from large data sets using a framework like Hadoop, HDFS, and Spark. Data sources used to create predictive models often include SQL databases, equipment log files, images, video, audio, and sensor data. These predictive models can then be deployed for production in an IT environment or on an embedded system.

Machine Learning Techniques

One of the key technologies enabling predictive analytics is machine learning. Techniques used in predictive analytics often include linear and nonlinear regression, neural networks, support vector machines, decision trees, and other machine learning algorithms.



Learn More

> See the ebook, *Machine Learning with MATLAB*

Your Data + MATLAB = Success with Predictive Analytics

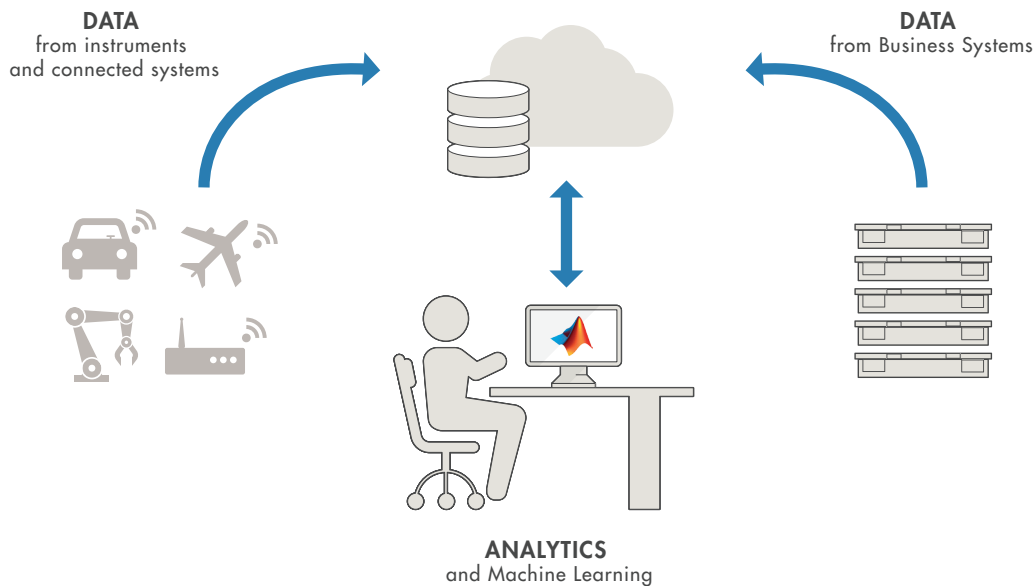


Figure 1. Architecture of engineering-driven analytics.

In this simplified view, engineering data arrives from sensors, instruments, and connected systems out in the world. The data is collected and stored in a file system either in-house or in the cloud—often in an aggregator of some kind.

“No matter what industry our client is in, and no matter what data they ask us to analyze—text, audio, images, or video—MATLAB code enables us to provide clear results faster.” — DR. G. SUBRAHAMANYA VRK RAO, COGNIZANT

This data is combined with data sourced from traditional business systems such as cost data, sales results, customer complaints, and marketing information. You now have a single store for all your data.

This is where the analytics are applied. Analytical methods such as statistics, machine learning, and signal processing are used against the combined data store, to produce an “analytic” —a predictive model of your system.

To be useful, that predictive model is then deployed—either in a production IT environment feeding a real-time transactional or IT system such as an e-commerce site or to an embedded device—a sensor, a controller, or a smart system in the real-world such as an autonomous vehicle.

Applying MATLAB and Simulink® together on top of this architecture is ideal, because the tools integrate directly into your embedded systems with Model-Based Design or IT system workflows. But more importantly, you can deploy to both.

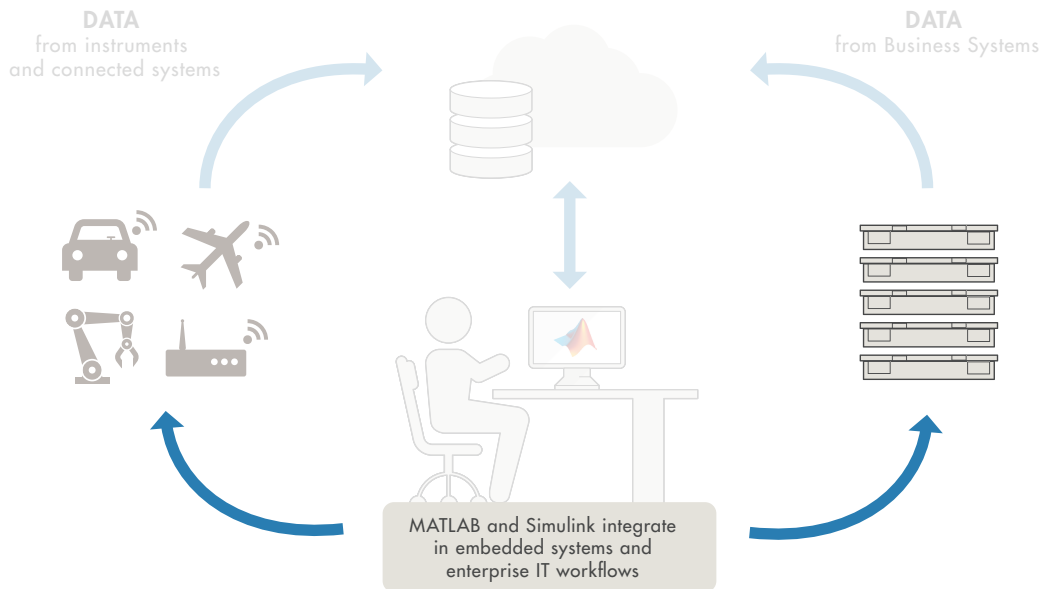


Figure 2. Deploying predictive models to embedded systems and IT systems.

Industry Examples of Predictive Analytics

Business and engineering teams in vastly different industries are using MATLAB and Simulink to solve big problems with predictive analytics.

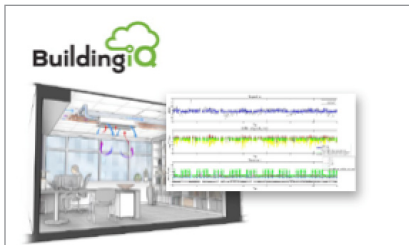
Predictive analytics helps teams in industries as diverse as finance, healthcare, pharmaceuticals, automotive, aerospace, machinery, and more.

“The value of looking at data analytics is an estimated at 6-8% uplift in production, which is significant. We are talking in the billions.” — AMJAD CHAUDRY, SHELL OIL

> [Watch video](#)

BuildingIQ – Online Optimization of Building Energy Use

BuildingIQ used MATLAB to speed the development and deployment of its predictive energy optimization algorithms. The application uses MATLAB to analyze and visualize big data sets, implement advanced optimization algorithms, and run the algorithms in a production cloud environment.



Online optimization of building energy use:

- Real-time, cloud-based system for commercial building owners to reduce energy consumption of HVAC operation
- Combines analytics and machine learning with optimization for predictive control of single-building HVAC
- Energy consumption reduced 15-25%

> [Read the user story](#)

Safran Snecma – Online Engine Health Monitoring

To monitor engines, Snecma has created an environment for the design and development of health monitoring algorithms: SAMANTA (Snecma Algorithm AND Test Application). This platform, used with MATLAB and Simulink, enables users to integrate algorithmic applications such as word processing, input/output, or display, as well as to continuously and automatically optimize algorithms.



Application for online engine health monitoring:

- Real-time analytics integrated with enterprise service systems
- Predict sub-system performance (oil, fuel, liftoff, mechanical health, controls)
- Improve aircraft availability and reduce maintenance costs

> Watch video: [Presentation of a Platform for the Development of Aircraft Engine Monitoring Algorithms: SAMANTA](#)

Respiri – Asthma and COPD Management

Engineers at Respiri have developed technology that asthma patients can use to record and analyze their own breathing. Instead of listening for wheeze sounds, the AirSonea® technology detects wheeze patterns in an image created from recorded breath sounds. Respiri used MATLAB to develop acoustic respiratory monitoring algorithms, and used *MATLAB Coder*™ to implement them as a mobile app and cloud-based server software.



Embedded-system and cloud-based app for patient diagnosis:

- Medical device to monitor and manage asthma and COPD
- Leverages analytics in cloud and embedded system
- Invokes spectral processing and pattern-detection analytics for wheeze detection on Respi server in the cloud
- Provides feedback to patients on their smart phones

“MATLAB enables us to rapidly develop, debug, and test sound-processing algorithms, and MATLAB Coder simplifies the process of implementing those algorithms in C. There’s no other environment or programming language that we could use to produce similar results in the same amount of time.” —YULYA GORYACHEV, RESPIRI

[> Read the user story](#)

Mondi Gronau

Mondi Gronau’s plastic production plant delivers about 18 million tons of plastic and thin film products annually. Machine failures that result in downtime and wasted raw materials cost Mondi millions of euros each month. To minimize these costs and maximize plant efficiency, Mondi developed a health monitoring and predictive maintenance application. The application uses advanced statistics and machine learning algorithms to identify potential issues with the machines, enabling workers to take corrective action and prevent serious problems.



Reducing costs with predictive maintenance for industrial equipment:

- Use MATLAB to develop and deploy monitoring and predictive maintenance software that uses machine learning algorithms to predict machine failures
- 50,000 Euros saved monthly
- Prototype completed in six months
- Production software running 24/7

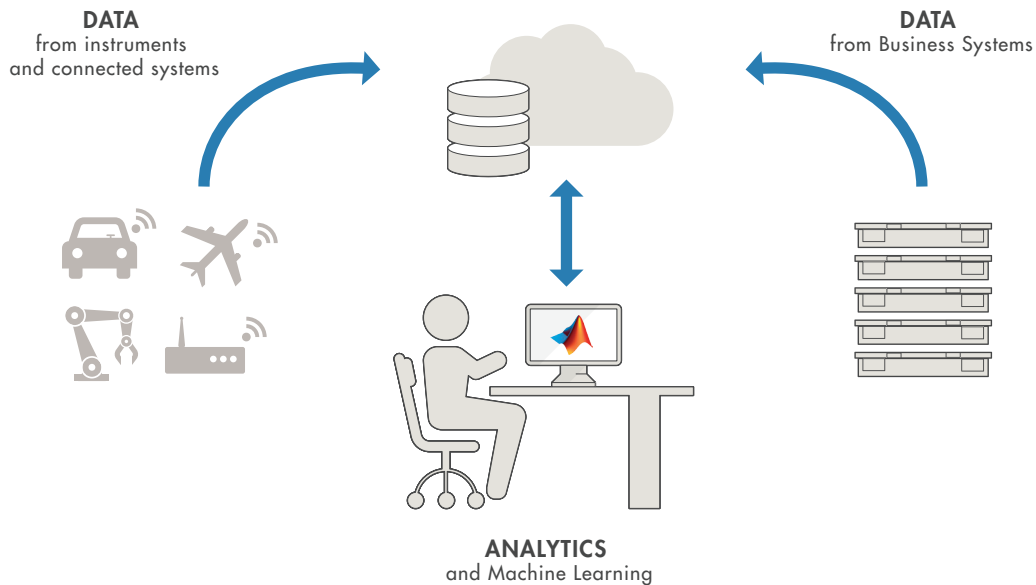
“As a manufacturing company, we don’t have data scientists with machine learning expertise, but MathWorks provided the tools and technical knowhow that enabled us to develop a production preventative maintenance system in a matter of months.”

—DR. MICHAEL KOHLERT, HEAD OF INFORMATION MANAGEMENT AND PROCESS AUTOMATION, MONDI

[> Read the user story](#)

MATLAB Users Make Great Data Scientists

Thinking again of the engineering-driven analytics architecture, what about the human in the middle?



Who will do the analysis and build the analytics? The answer is a data scientist. Data scientists combine three types of knowledge.

First, they have domain expertise. They are experts in the field in which they work. They know the engineering or the science behind their projects. Second, they understand computing. They know the basics of coding, data management, and computing infrastructure. Third, they understand how to apply statistical and mathematical analysis to their problems.

It is very rare to find people who have all three types of knowledge, so that makes them difficult to hire. Almost every week there are articles written about the shortage of data scientists, and it is recognized as a global issue.

Rather than hiring costly data scientists who may not have the correct domain expertise, you can look to MATLAB users to apply statistical methods and machine learning to solve your engineering problems. These engineers and scientists already have the domain expertise, so they are able to quickly determine whether an analytic technique is going to be useful.

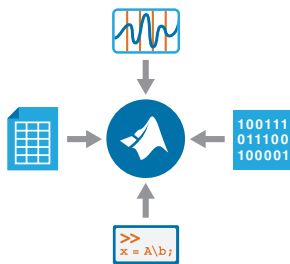
Finding Better Solutions, Faster, with MATLAB

Using MATLAB, your teams can try out more ideas, which leads to more effective and efficient results.

Focus your team's efforts in these four steps:

1. Access and explore your data.

MATLAB is designed to handle large amounts of data efficiently, with native support for physical-world data in sensor, image, video, telemetry, binary, and other real-time formats.



“One data set captured 100,000 hours of sound. It was so large that we had previously processed less than 1% of it, estimating that it would take a year or more to process the rest. With our MATLAB high-performance computing platform, we processed the data six times, using different detection algorithms, in two days.”

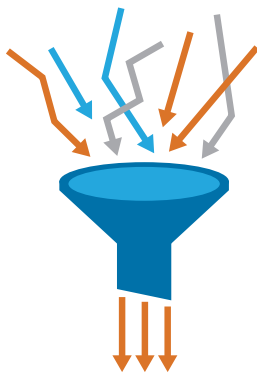
—PETER DUGAN, LEAD DATA SCIENTIST, BIOACOUSTICS RESEARCH PROGRAM CORNELL LABORATORY OF ORNITHOLOGY

2. Preprocess your data.

MATLAB provides tools to make it faster and easier to do high-speed processing of large data sets. In addition, its numeric routines scale directly to parallel processing on clusters and cloud.

The MATLAB environment provides preprocessing techniques such as advanced signal processing, for removing noise from sensor data; image processing, for sharpening an image and isolating objects of interest; and feature selection, extraction, and transformation, for reducing the dimensions of your data sets.

With MATLAB you can do your thinking and programming in one environment.



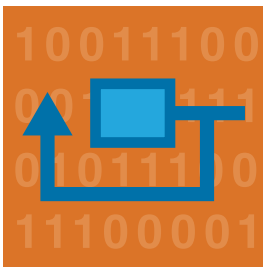
“We need to filter our data, look at poles and zeroes, run nonlinear optimizations, and perform numerous other tasks. In MATLAB, those capabilities are all integrated, robust, and commercially validated.”

—BORISLAV SAVKOVIC, LEAD DATA SCIENTIST, BUILDINGIQ

3. Develop and iteratively test predictive models on your data.

Your aggregated data tells a complex story. To extract the insights it holds, you need an accurate predictive model. You can quickly select and identify the right features for a model and then iterate through additional models to identify the best algorithm.

MATLAB speeds up this process with specialized toolboxes, prebuilt functions, and a full set of *statistics and machine learning functionality*. It also offers advanced methods such as nonlinear optimization, system identification, and thousands of prebuilt algorithms for image and video processing, financial modeling, control system design.



“Because our team already knew MATLAB, we did not need a programmer. Instead, our structuring and market operations analysts, who already had the necessary experience in mathematics and economics, developed the system. MATLAB enabled these analysts to build a reliable, scalable forecasting and analysis solution from scratch.”

—MANUEL ARANCIBIA, MARKETING OPERATIONS MANAGER,
HORIZON WIND ENERGY

4. Integrate analytics with your production systems and target to existing IT systems and embedded hardware.

MATLAB offers online and real-time deployment that integrates into enterprise systems, clusters, and clouds, and can be targeted to real-time embedded hardware.



“The tools we developed with MATLAB are more reliable, scalable, and maintainable than our spreadsheet-based approach. We know the tools will work, we can add new capabilities, and we can update the production system without getting IT involved.”

—MANUEL ARANCIBIA, MARKETING OPERATIONS MANAGER,
HORIZON WIND ENERGY

Conclusion

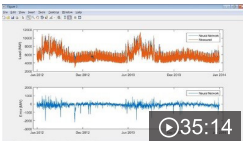
Predictive analytics enables informed decision making in the face of uncertainty and risk, through data-driven insights and predictions.

MATLAB provides an interactive environment and a powerful set of tools to enable domain experts to become data scientists—developing custom predictive models from engineering and business data. Through flexible deployment options, production-ready models can be integrated much more quickly into business systems and embedded devices.

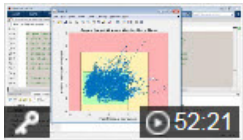
Learn More



Big Data and Predictive Analytics at Shell (Video)



Data Analytics with MATLAB (Webinar)



*Predictive Maintenance with MATLAB:
A Prognostics Case Study (Webinar)*

Free product trial – *Try MATLAB for Data Analytics today*