

Machine Learning Made Easy: InterSystems IntegratedML

Technology Brief





Win the Artificial Intelligence Talent War With an Easy-to-Develop, Easy-to-Deploy Machine Learning Solution

Why Read this Technology Brief?

According to Forrester Research, “98% of companies experience challenges with gaining insights from the data they collect; this is primarily due to the lack of internal expertise.”¹ While your organization may be able to maintain its competitive edge today without these insights, the pace of change toward digital transformation may soon affect your business.

Hence, organizations everywhere are focused on using data – and incorporating artificial intelligence (AI) and machine learning (ML) – to improve their businesses. With ML, you can improve and automate business operations, predict events and behaviors, and proactively execute prescriptive programmatic actions based on these predictions.

For example, using ML and predictive analytics, your organization can identify and target new buyers or identify the best times to run a sale by better understanding consumer behavior and preferences. If you are a healthcare provider, your organization can analyze coded diagnoses as well as the patient’s admission, transfer, and discharge data to lower re-admission rates. Simply put, ML can save time and resources, improve forecasting, and enable you to make better decisions and realize better outcomes.

It sounds great but there is one drawback: Developing ML models is difficult and requires scarce – and expensive – expertise.

Fortunately, industry experts are working to make ML easier to use by developing new tools, including AutoML and IntegratedML.

By reading this technology brief, you will discover:

- Why ML is critical to your business’ success
- How a talent shortage is making it challenging for organizations to leverage ML
- What AutoML is and how it helps you win the AI talent war
- What InterSystems IntegratedML™ is and how it is designed to:
 - Empower your existing software developers to develop ML models and ML-enabled applications.
 - Increase the productivity of trained data scientists.
 - Streamline operational and analytical processes to improve customer experience, operational efficiency, and productivity.
 - Improve forecast accuracy, create better business outcomes, and enable you to differentiate from your competitors.

¹ Forrester Opportunity Snapshot. (2019) Data Insights Are Key to Differentiated Customer Experience: A Unified Data Analytics Platform Enables Timely and Contextually Relevant CX

This technology brief is appropriate reading for line of business executives, managers, and IT professionals, whether you are looking to extend the productivity of your ML team or just getting started with ML without the need to hire ML experts.

Machine Learning: The Value

As an application of AI, ML trains a machine to learn about data from experience and inference. It continuously improves outcomes without being specifically programmed to do so.

ML can analyze a wide range of data and create models used to meet a vast array of analytic and operational requirements. Offline, ML models can help business users understand customer behavior, or process efficiency problems, to name just a few applications. When deployed online, or in the operational flow of a business, ML can very visibly deliver improved outcomes – whether that’s recommending a preferred product or service to a customer while she is browsing, proactively alerting you before making a sale if there is a high risk the supplier will be unable to deliver, or determining whether a transaction may be fraudulent before approving it. Department operations in every part of your organization can benefit from ML, including sales and marketing, research and development, legal, human resources, customer support, product development, even finance. The fact is, ML is providing value in almost every industry, and promises to become ubiquitous as more and more organizations embrace it.

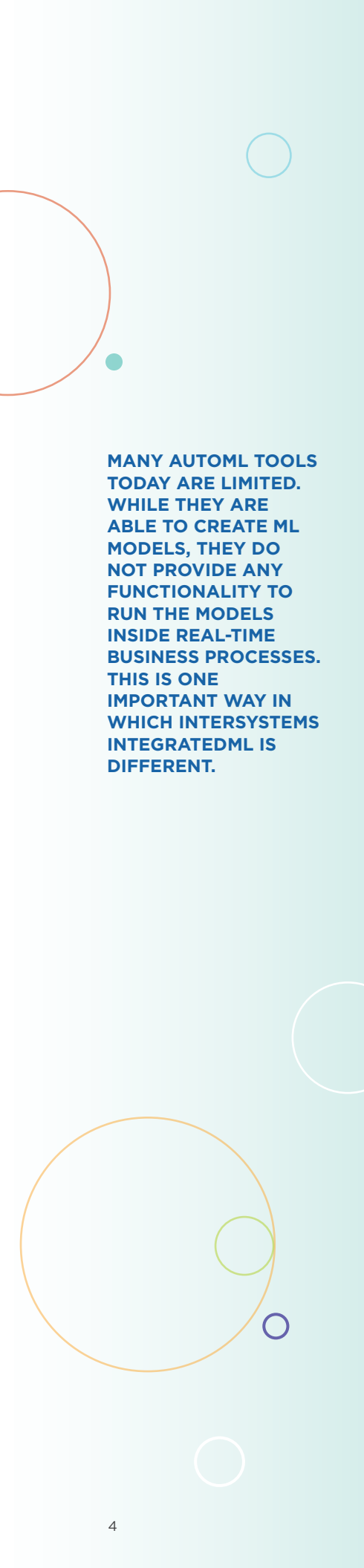
You already are experiencing ML in your everyday life: from virtual personal assistants such as Amazon Alexa and Apple’s Siri, to spam filters and malware detectors, to Facebook’s method for suggesting new friends and new groups, to chatbots that provide online customer support, to smart cars that drive themselves.

Machine Learning: The Challenge

Machine learning offers up many benefits, but it begs the question: why aren’t more companies using it? One key reason: ML is difficult to use and requires a high level of expertise.

ML requires experts who understand the theory, technology, methods, and tools. Today, these experts are few and far between and in high demand. According to the latest data from the U.S. Bureau of Labor Statistics, there are less than 32,000 data scientists in total in the U.S.² Compounding the shortage of AI specialists and data scientists, much of the talent that is available is being hired by the likes of the digital giants, including Amazon, Facebook, Google, and Microsoft, who are paying dizzying high salaries. This makes it difficult for organizations to compete for these already scarce resources.

² <https://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm>



MANY AUTOML TOOLS TODAY ARE LIMITED. WHILE THEY ARE ABLE TO CREATE ML MODELS, THEY DO NOT PROVIDE ANY FUNCTIONALITY TO RUN THE MODELS INSIDE REAL-TIME BUSINESS PROCESSES. THIS IS ONE IMPORTANT WAY IN WHICH INTERSYSTEMS INTEGRATEDML IS DIFFERENT.

AutoML: Winning the AI Talent War

Automated Machine Learning (AutoML) is a burgeoning new technology for organizations looking to expand the reach of their current ML talent and for those that are just starting on their ML journey.

AutoML is a relatively new approach to data science – it automates and simplifies the creation of ML models. It performs feature engineering, automating the process of transforming raw data into formats appropriate for ML models. It automates model selection, training and results analysis – and tests different ML algorithms with varying parameters to create the most accurate model for any given problem. For organizations with a team of data scientists, this automates much of the manual and trial-and-error processes used to build ML models and significantly improves the productivity of your data scientists, saving time and effort.

If you don't already have ML specialists on staff, performing feature engineering and creating and training models can be challenging. But now, with AutoML, your organization doesn't necessarily need data scientists to create useful ML models. Instead, you can start with simple use cases and AutoML, while simultaneously training your developers to undertake more of the analysis and ML development process.

However, many AutoML tools today are limited. While they are able to create ML models, they do not provide any functionality to run the models inside real-time business processes. This is one important way in which InterSystems IntegratedML is different.

InterSystems IntegratedML: AutoML to a Higher Power

InterSystems IntegratedML is an embedded feature of the InterSystems IRIS® data platform, a complete data management software environment. IntegratedML provides all of the features and benefits of traditional AutoML. Since it is embedded within InterSystems IRIS, however, you can develop and deploy sophisticated applications that seamlessly execute these models dynamically in response to real-time events and transactions, without extracting or moving any models or data.

For example, consider a bank that issues credit cards that needs to identify fraud risk before approving each transaction. It executes a real-time, high-performance credit card application developed with InterSystems IRIS, which stores all of the demographic and financial data of all customers and credit card transactions. This application can contain hundreds of data elements for each credit card transaction – including whether each transaction was fraudulent or valid.

Using IntegratedML, the existing application developers at the bank can automatically create an ML model to identify high-risk transactions based on past transactions, by simply selecting the desired field (e.g. “is_fraudulent”) and letting IntegratedML create the most appropriate model and parameters.

But unlike traditional AutoML, the InterSystems IntegratedML-based model can be seamlessly incorporated into the credit card application to execute in real time with each incoming transaction, and the application can take the appropriate programmatic actions if the model determines that there is a high risk of fraud, such as preventing the transaction and calling and texting the card owner.

IntegratedML also makes it easier to keep models current as applications run in production and new data is generated. In the case of credit card fraud, as one mode of fraud is detected and prevented by the application, new techniques will surely be initiated by the criminals. Since all the data, including the most recent data, is stored within the data platform, there is no need to create manual extracts and move data to different environments. Instead, the bank can continuously refine the models using the most recent data to detect and prevent new attack patterns, without delay.

With InterSystems IRIS and IntegratedML, you can develop applications that perform intelligent prescriptive programmatic actions in response to real-time events and gain critical competitive advantages and business benefits. It can help you be first to market with a new product or service, first to act on a new initiative, and first to respond to a change in customer behavior.

IntegratedML: Reduce Talent Costs and Improve Productivity

With IntegratedML, a developer – with little to no ML knowledge – can use SQL to develop sophisticated ML models.

This is not to suggest that you should never hire data scientists. If your organization is a large enterprise with a team of data scientists, IntegratedML can save your data engineers and data scientists significant time. For example, a 2018 survey conducted by [Kaggle ML and Data Science](https://www.kaggle.com/headsortails/what-we-do-in-the-kernels-a-kaggle-survey-story)³ found that data scientists spend almost 40% of their time gathering and cleaning data⁴ (see Figure 1 below). Using IntegratedML for data preparation and feature engineering can free up your data scientists to focus on more important, higher-value tasks such as optimizing models.

**WITH INTEGRATEDML,
A DEVELOPER – WITH
LITTLE TO NO ML
KNOWLEDGE – CAN
USE SQL TO DEVELOP
SOPHISTICATED ML
MODELS. IF YOUR
ORGANIZATION
ALREADY HAS A TEAM
OF DATA SCIENTISTS,
INTEGRATEDML
CAN SAVE THEM
SIGNIFICANT TIME
AND EFFORT.**

³ <https://www.kaggle.com/headsortails/what-we-do-in-the-kernels-a-kaggle-survey-story>

⁴ <https://businessoverbroadway.com/2019/02/19/how-do-data-professionals-spend-their-time-on-data-science-projects/>

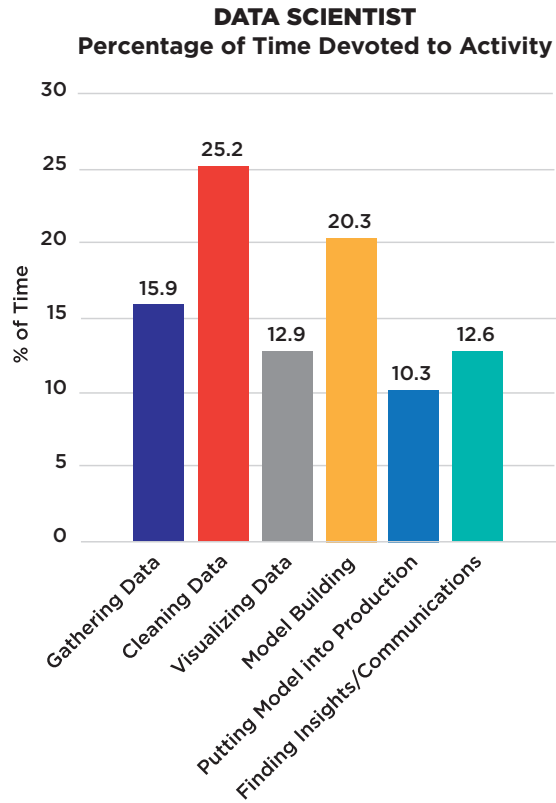


Figure 1: How Data Scientists Spend Their Time³

For organizations just getting started with ML, InterSystems IntegratedML lets the software developers and analysts who are building your business applications and know the data explore ML on their own. IntegratedML automates the basic work, such as identifying the most appropriate models, setting parameters, and building and training models. It also speeds the process of integrating the ML models into production applications. As your developers become more sophisticated and begin to understand the process and results, they can start to modify optional parameters and set the values themselves. Data scientists can also be more productive with IntegratedML because they can spend their time on the actual model optimization rather than wrangling data and feature engineering and selection.

InterSystems Integrated ML: How it Works

With IntegratedML, model training, including identifying proper input features from source data, tuning model parameters, and execution are all accomplished through the use of just a handful of SQL commands.

```
CREATE MODEL WillSurvive PREDICTING (Survived) FROM Titanic
```

The CREATE MODEL command sets up the machine learning model metadata. Developers specify the name of the model (WillSurvive), the target field to be predicted (Survived) and a dataset to source the target field and all model input fields from (Titanic). The FROM syntax is fully general and can specify any subquery expression. The metadata associated with this dataset is also used to infer the data types of the target and input fields, fully defining the problem for the model to solve.

```
TRAIN MODEL WillSurvive FROM Titanic
```


The TRAIN MODEL command specifies the data to be used for training and executes the AutoML engine, which takes as input a set of relational data. Since the FROM syntax is general, the same model can be trained multiple times with different sets of data. For example, you may want to train a marketing campaign model on different customer segments, or re-train your model on a regular basis, as new training data becomes available.

The AutoML engine automatically takes care of all required machine learning tasks. It identifies relevant candidate features from the selected data, considers applicable model types based on the data and problem definition, and tunes the hyper-parameters to yield one or more runnable models.

Developers can choose from different AutoML engines including InterSystems AutoML, H2O, and DataRobot Enterprise AI Platform. All AutoML engine options are seamlessly integrated within InterSystems IRIS and are transparent to developers.

```
SELECT PREDICT(WillSurvive) As Predicted FROM Titanic  
SELECT PROBABILITY(WillSurvive FOR 1) FROM Titanic
```

**DEVELOPERS CAN
CHOOSE FROM
DIFFERENT AUTOML
ENGINES.**



Once trained, the model provides results via one of two scalar functions, `PREDICT()` and `PROBABILITY()`. `PREDICT()` returns the most likely or estimated value for the specified column as determined by the trained model. For categorization problems, `PROBABILITY()` returns the trained model's calculated probability that the model's target field will be equal to a user defined value. These simple scalar functions can be used anywhere in a query and in any combination with other fields and functions. One of the key innovations IntegratedML provides is transparently taking care of mapping the available fields in the given query context to the input fields required to execute the model.

IntegratedML provides additional flexibility for developers, for example to map to other data sources than the particular table or query used to create or train the model, as illustrated by the following example.

```
SELECT Name, PREDICT(WillSurvive WITH Sex = Geschlecht,  
Age = DATEDIFF(year, NOW(), Geburtsdatum), Fare =  
TicketPreise, Cabin = Kabine) FROM Hindenburg
```

While most AutoML solutions operate in a standalone environment with loose, low-throughput coupling with external data platforms and applications, IntegratedML is different. It operates seamlessly within the InterSystems IRIS data platform to speed and simplify the training and execution of ML models, and allows the ML models to be seamlessly integrated within InterSystems IRIS applications without moving the data or models. This operationalization of ML models is considered to be one of the biggest impediments to swift adoption of ML in business applications.

InterSystems IRIS Data Platform

InterSystems technology powers 150,000 deployments worldwide across a variety of industries. InterSystems IRIS data platform is a complete data management software platform purpose built to speed and simplify the development of real-time, data-driven applications. InterSystems IRIS allows developers to incorporate sophisticated analytics — including business intelligence, AI, ML, natural language processing, and predictive analytics — into real-time, mission-critical business processes. The embedded high performance transactional-analytic database engine concurrently supports both operational and analytic workloads at very high scale.

In addition to its embedded ML development and runtime capabilities, InterSystems IRIS also enables:

Data and Application Integration – InterSystems IRIS provides a complete set of integration and interoperability capabilities to clean, transform, and normalize data, and support sophisticated integrations. It provides out-of-the-box connectivity and data transformations for a wide range of packaged applications, databases, industry standards, protocols, and technologies to make it easier to integrate and analyze data and build predictive and prescriptive models.

In addition, you can embed analytic processing, such as SQL queries, predictive analytics, ML, and Natural Language Processing (NLP) into composite business processes that connect disparate data sources and applications. These composite processes can streamline operations, trigger alerts, and do so without impacting application performance.

Scalability – InterSystems IRIS is vertically and horizontally scalable, and highly resource efficient, making it ideal for applications that support very high-volume ingestion rates, high levels of analytic workloads, many concurrent business processes, and the ability to process, store, and analyze very large data sets in a cost-effective manner.

Reporting and Traceability – All data (including in-flight data, metadata, and data associated with long-running asynchronous transactions) is automatically stored in the embedded database and available for real-time reporting and analysis. Visualizing and diagnosing the behavior of integrations and processes is made easier through visual trace capabilities.

Graphical Development – Graphical, low-code tooling allows developers to visually diagram processes, transformations, rules, and workflows, so that they can focus on the logical interactions between systems rather than coding. The graphical models encourage collaboration between the lines of business and IT, allowing your organization to develop new solutions or modify existing applications faster.

Deployment – InterSystems IRIS supports a wide range of deployment options, including all major public clouds, private clouds, on premises, and hybrid deployment options.

Whether you are looking to delight your customers with real-time personalized experiences, improve clinical outcomes for patients, proactively predict maintenance needs in advance of failures, or detect and prevent fraud in real time, InterSystems IRIS and IntegratedML can help you achieve these objectives and more.



Conclusion

Machine learning is the wave of the future and any organization looking to compete will need to start using it. Unfortunately, data scientists are scarce and their salaries are skyrocketing, making it challenging for large organizations to expand their ML footprint and small organizations to get started with ML. While innovations such as AutoML are helping, AutoML alone is not sufficient.

InterSystems IntegratedML provides sophisticated AutoML capabilities, exposed through an intuitive SQL interface, and fully integrated within a comprehensive data platform. IntegratedML makes it easy to deploy ML models in real-time, mission critical applications without the need to move data or models, and without requiring a staff of data scientists. Together, InterSystems IRIS and IntegratedML enable you to create a virtuous cycle of improvement, continually refining ML models without delay in response to the most recent production data.

If you have a team of data scientists, IntegratedML will improve your team's productivity.

If you are just starting out on your AI journey, IntegratedML can get you started with ML now, without hiring expensive ML experts.

In either case, IntegratedML can help you:

- Speed and simplify the creation of ML models
- Execute intelligent programmatic actions in real time
- Streamline processes to improve customer experiences, operational efficiency, and productivity
- Improve forecast accuracy, accelerate better business outcomes, and outmaneuver your competition
- Develop smarter apps faster and easier with fewer resources
- Win the AI talent war

For more information about InterSystems IRIS and IntegratedML, visit [InterSystems.com/IRIS](https://interSystems.com/IRIS)

About InterSystems

InterSystems is the information engine that powers some of the world's most important applications. In healthcare, finance, government, and other sectors where lives and livelihoods are at stake, InterSystems has been a strategic technology provider since 1978. InterSystems is a privately held company headquartered in Cambridge, Massachusetts (U.S.), with offices worldwide, and its software products are used daily by millions of people in more than 80 countries. For more information, visit [InterSystems.com](https://www.intersystems.com).

