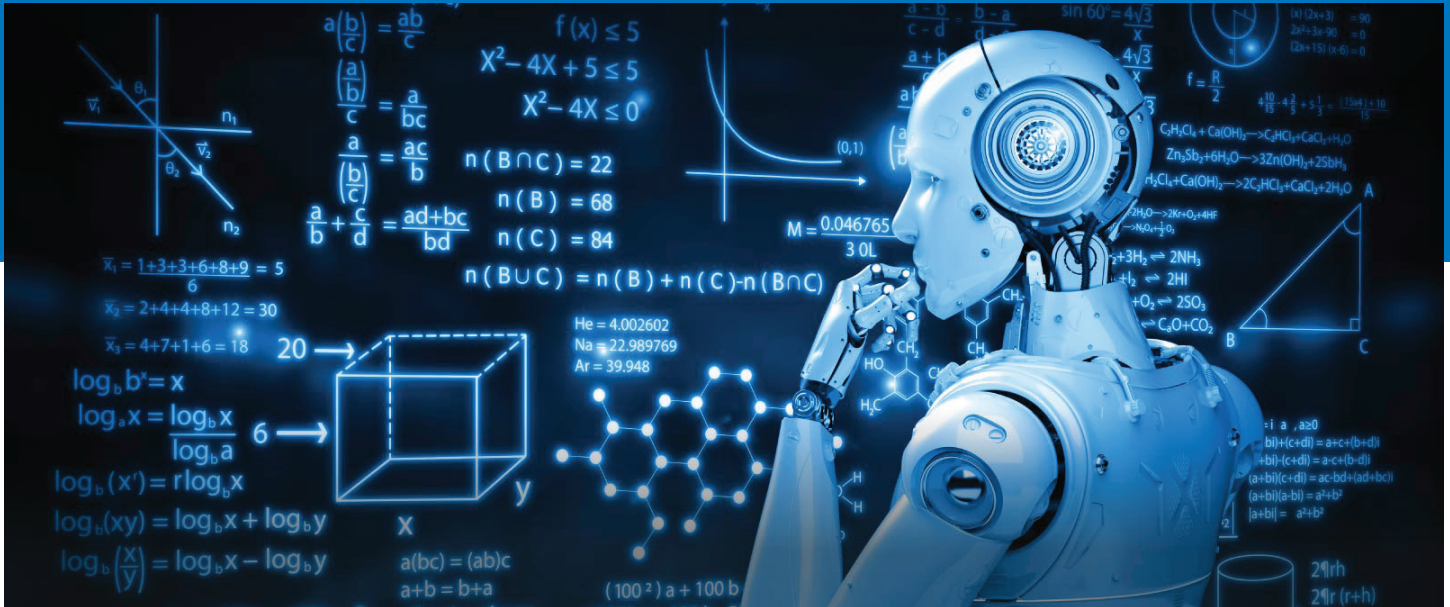# Machine Learning Supercharges Digital Twins for Real-Time Analytics

By Dr. William Bain, Founder & CEO | ScaleOut Software

HIGHLIGHTS

▶ "Real-time" digital twins combine the digital twin model with in-memory computing to perform streaming analytics at blazingly fast speeds and scale.

▶ ScaleOut Software's real-time digital twins can use machine learning to automatically detect important anomalies in streaming data.

▶ Using the ScaleOut Model Development Tool, developers can easily incorporate Microsoft's ML.NET machine learning library into their digital twins.

REAL-TIME DIGITAL TWINS ENABLE FAST STREAMING ANALYTICS

When tracking telemetry from a large number of IoT devices, it's essential to quickly detect when something goes wrong. For example, a fleet of long-haul trucks needs to meet demanding schedules and can't afford unexpected breakdowns as a fleet manager manages thousands of trucks on the road. With today's IoT technology, these trucks can report their engine and cargo status every few seconds to cloud-hosted telematics software. How can this software sift through the flood of incoming messages to identify emerging issues and avoid costly failures? Can the power of machine learning be harnessed to provide predictive analytics that automates the task of finding problems that are otherwise very difficult to detect?

ScaleOut Software

Real-time digital twins offer a powerful software architecture for tracking and analyzing IoT telemetry from large numbers of data sources. A real-time digital twin is a software component running within a fast, scalable in-memory computing platform, and it hosts analytics code and state information required to track a single data source, like a truck within a fleet. Thousands of real-time digital twins run together to track all of the data sources and enable highly granular real-time analysis of incoming telemetry. By building on the widely used digital twin concept, real-time digital twins simultaneously enhance real-time streaming analytics and simplify application design.

MACHINE LEARNING WITH MICROSOFT'S ML.NET

Incorporating machine learning techniques into real-time digital twins takes their power and simplicity to the next level. While analytics code can be written in popular programming languages, such as Java and C#, or even using a simplified rules engine, creating algorithms that ferret out emerging issues hidden within a stream of telemetry still can be challenging. In many cases, the algorithm itself may be unknown because the underlying processes which lead to device failures are not well understood. In these cases, a machine learning (ML) algorithm can be trained to recognize abnormal telemetry patterns by feeding it thousands of historic telemetry messages that have been classified as normal or abnormal. No manual analytics coding is required. After training and testing, the ML algorithm can then be put to work monitoring incoming telemetry and alerting when it observes suspected abnormal telemetry.

To enable ML algorithms to run within real-time digital twins, ScaleOut Software has integrated Microsoft's popular machine learning library called ML.NET into its Azure-based ScaleOut Digital Twin Streaming Service™. Using the ScaleOut Model Development Tool™ (formerly called the ScaleOut Rules Engine Development Tool), users can select, train, evaluate, deploy, and test ML algorithms within their real-time digital twin models. Once deployed, the ML algorithm runs independently for each data source, examining incoming telemetry within milliseconds after it arrives and logging abnormal events. The real-time digital twin also can be configured to generate alerts and send them to popular alerting providers, such as Splunk, Slack, and Pager Duty. In addition, business rules optionally can be used to further extend real-time analytics.

The following diagram illustrates the use of an ML algorithm to track engine and cargo parameters being monitored by a real-time digital twin hosting an ML algorithm for each truck in a fleet. When abnormal parameters are detected by

the ML algorithm (as illustrated by the spike in the telemetry), the real-time digital twin records the incident and sends a message to the alerting provider:
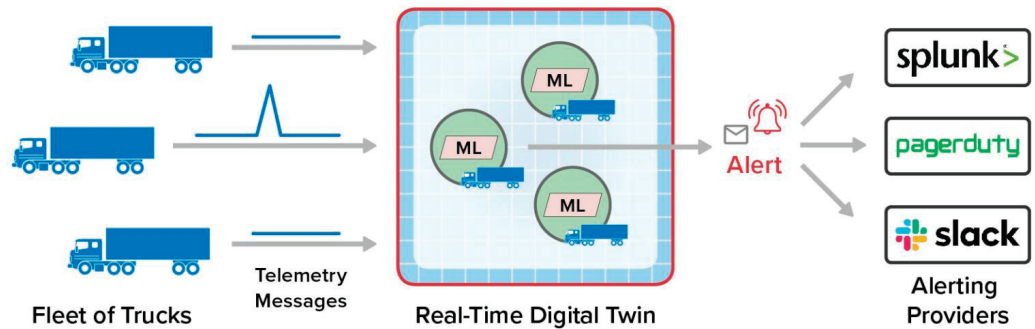
USING THE SCALEOUT MODEL DEVELOPMENT TOOL

Training an ML algorithm to recognize abnormal telemetry just requires supplying a training set of historic data that has been classified as normal or abnormal. Using this training data, the ScaleOut Model Development Tool lets the user train and evaluate up to ten binary classification algorithms supplied by ML.NET using a technique called supervised learning. The user can then select the appropriate trained algorithm to deploy based on metrics for each algorithm generated during training and testing. (The algorithms are tested using a portion of the data supplied for training.)

For example, consider an electric motor which periodically supplies three parameters (temperature, RPM, and voltage) to its real-time digital twin for monitoring by an ML algorithm to detect anomalies and generate alerts when they occur:
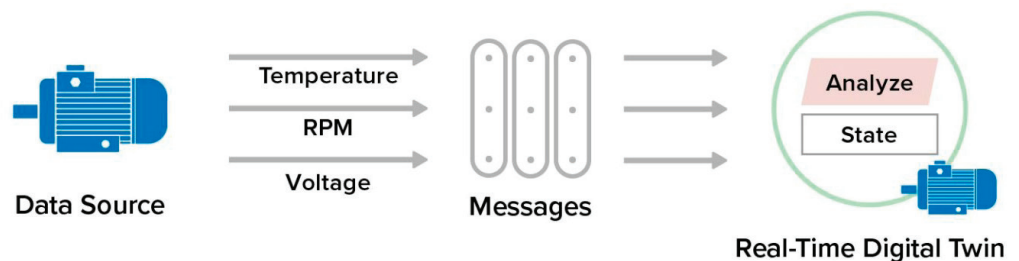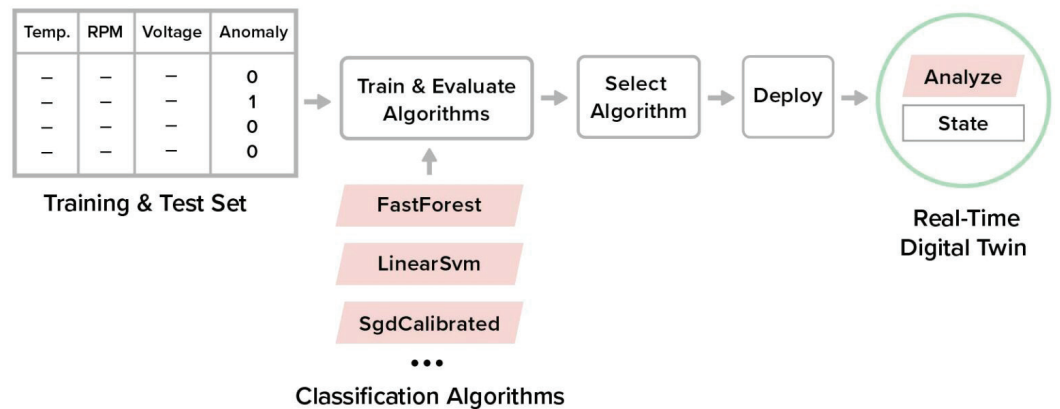

Anomaly Detection

Training the real-time digital twin's ML model follows the workflow illustrated below:

Workflow for Training Real-Time Digital Twin Model

Using supervised learning, users train an ML algorithm for deployment in a real-time digital twin.

Here's a screenshot of the ScaleOut Model Development Tool that shows the training of selected ML.NET algorithms for evaluation by the user:


The output of this process is a real-time digital twin model which can be deployed to the streaming service. As each motor reports its telemetry to the streaming service, a unique real-time digital twin "instance" (a software object) is created to track that motor's telemetry using the ML algorithm.

In addition to supervised learning, ML.NET provides an algorithm (called an adaptive kernel density estimation algorithm) for spike detection, which detects rapid changes in telemetry for a single parameter. The ScaleOut Model Development Tool lets users add spike detection for selected parameters using this algorithm. In addition, it is often useful to detect unusual but subtle changes in a parameter's telemetry over time. For example, if the temperature for an electric motor is expected to remain constant, it would be useful to detect a slow rise in temperature that might otherwise go unobserved. To address this need, the tool lets users make use of a ScaleOut-developed, linear regression algorithm that detects and reports inflection points in the telemetry for a single parameter. These two techniques for tracking changes in a telemetry parameter are illustrated below:

Spike Detection

Trend Detection
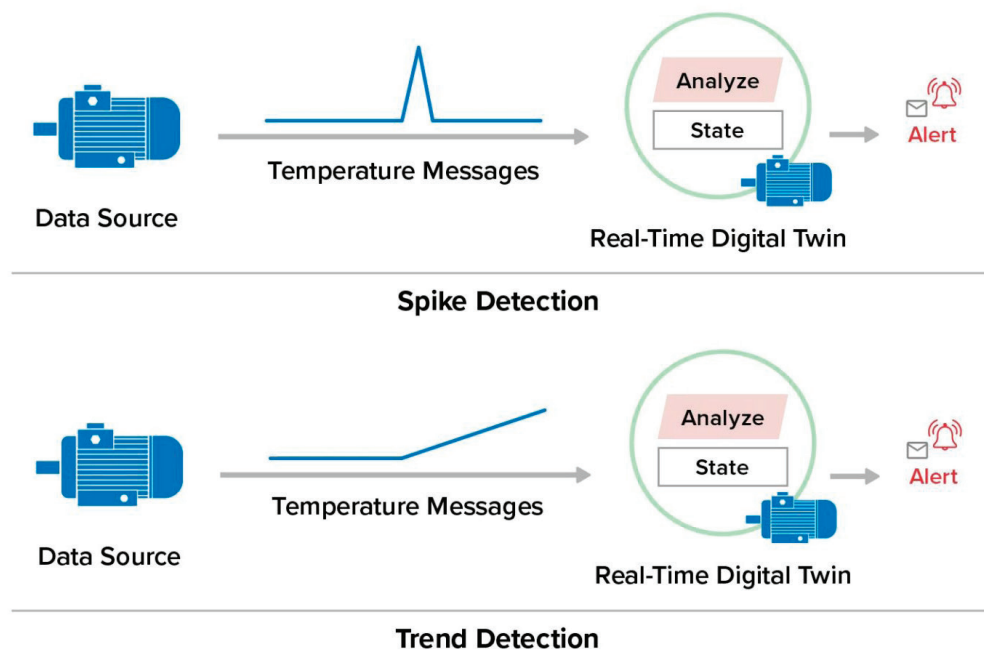
SUMMING UP

Machine learning provides important real-time insights that enhance situational awareness and enable fast, effective responses. They often can provide useful analytics for complex datasets that cannot be analyzed with hand-coded algorithms. Their usefulness and rate of adoption is quickly growing. Using the ScaleOut Model Development Tool, real-time digital twins now can easily be enhanced to automatically analyze incoming telemetry messages with machine learning techniques that take full advantage of Microsoft's ML.NET library. The integration of machine learning with real-time digital twins enables thousands of

data streams to be automatically and independently analyzed in real-time with fast, scalable performance. Best of all, no coding is required, enabling fast, easy model development. By combining ML with real-time digital twins, the ScaleOut Digital Twin Streaming Service adds important new capabilities for real-time streaming analytics that supercharge the Azure IoT ecosystem.

ABOUT THE AUTHOR

Dr. William L. Bain is founder & CEO of ScaleOut Software, Inc. and holds a Ph.D. in electrical engineering from Rice University. Bill has contributed to advancements in parallel computing at Bell Labs research, Intel and Microsoft, and holds several patents in computer architecture and distributed computing. Bill founded and ran three start-up companies prior to founding ScaleOut Software in 2003. His previous company, Valence Research, developed and distributed Web load-balancing software, and was acquired by Microsoft Corporation. As a member of the screening committee for the Seattle-based Alliance of Angels, Bill is actively involved in entrepreneurship and the angel community.

ABOUT SCALEOUT SOFTWARE

Founded in 2003, ScaleOut Software develops leading-edge software that delivers scalable, highly available, in-memory computing and streaming analytics technologies to a wide range of industries. ScaleOut Software's in-memory computing platform enables operational intelligence by storing, updating, and analyzing fast-changing, live data so that businesses can capture perishable opportunities before the moment is lost.

ScaleOut Software offers a full portfolio of in-memory computing products and technologies to deliver operational intelligence for your organization. Please visit www.scaleoutsoftware.com for more details on our technology and product portfolio for mission critical enterprise applications, including the ScaleOut Digital Twin Streaming Service™, an award-winning Azure-based cloud service and on-premises platform for hosting streaming analytics using the digital twin model. In addition, ScaleOut Software offers ScaleOut StateServer®, a linearly scalable, highly available in-memory data grid and ScaleOut StateServer® Pro, a data-parallel compute engine for blazingly fast, data-parallel computation on live, fast-changing data.

WINNER OF THE
2022 IoT ANALYTICS INNOVATION AWARD