

Automated Machine Learning

The next generation of IIoT for Predictive Maintenance



The data science discipline is rapidly evolving and there has been a shift towards the automation of certain repetitive and time-consuming machine learning tasks. This is referred to as Automated Machine Learning or AutoML.

As Rich Caruana, a senior researcher at Microsoft, noted at an [AutoML workshop](#), 75% of Machine Learning work involves preparing the data for Machine Learning models. In simple terms, AutoML is the

application of the science of Machine Learning to the practice of Machine Learning to reduce many of the repetitive tasks and improve algorithms' performance and accuracy.

This article will explain the meaning and benefit of AutoML and explore both its technical feasibility and relevance to the field of Industrial IoT (IIoT) or data driven Predictive Maintenance, also referred to as PDM4.0.

The background of AutoML

In recent years, the topic of AutoML has achieved prominence in the data science community. As Machine Learning is applied to more industrial and commercial applications, a need exists to replace scarce and highly skilled human Machine Learning expertise with automated systems. Google's announcement of their own proprietary AutoML tool, AutoML-Zero, in March 2020 is in itself indicative of AutoML's increasing importance within the discipline of Machine Learning.

Machine Learning is a difficult science requiring a high level of discipline expertise. What makes Machine Learning so challenging? In each step of the process, data scientists make choices that will impact the results or outcomes. There is no manual for the labour-intensive decisions of model selection and configuration, hyperparameter optimization, etc.

Let's look at the Machine Learning workflow using the example of IIoT for Predictive Asset Maintenance.

AutoML is applied primarily to three processes within the Machine Learning Pipeline.

How does AutoML work? AutoML is a continuous process that incorporates feedback from the data scientist on the pipelines that have been used to date. It adapts the optimization process of the Machine Learning Pipeline in real time and provides recommendations for model, process and hyperparameter selections.

According to Randy Olson, a leading data scientist and developer of the TPOT tool, "in the near future, I see automated machine learning (AutoML) taking over the machine learning model-building process: once a data set is in a (relatively) clean format, the AutoML system will be able to design and optimize a machine learning pipeline faster than 99% of the humans out there."

AutoML process optimization

Process	High-level description	Application of AutoML
Data pre-processing	Data is processed, cleaned, formatted and transformed before the Machine Learning Model can be trained on it.	Automatically chooses pre-processing model.
Model selection	The Machine Learning Model is selected from the dozens of possible popular Machine Learning models per problem type.	Automatically chooses Machine Learning model for a specific dataset according to some performance metric (accuracy, time or memory consumption).
Hyperparameter optimization	A set of optimal Model hyperparameters (e.g., different learning rates, constraints, weights, etc.) is selected for a Machine Learning Model from among thousands of possible configurations.	Automatically chooses a possible Model hyperparameter configuration from among thousands of possible configurations in order to maximize results.

It should be noted that it would be an oversimplification to view AutoML only within the context of efficiency and workflow automation. An equally important benefit is that it improves the accuracy and quality of the Machine Learning model. As senior data scientists Hamel Husain and Nick Handel wrote in KDnuggets, with AutoML the model is built rapidly, enabling data leakage to be detected early in the modelling lifecycle.

AutoML is still in its infancy and end users cannot use the various tools that have been released. At present, no one common definition of AutoML exists, and it has not been applied to all stages of the Machine Learning pipeline.

In the future, we expect more research into AutoML and the exploration of new areas for automation, such as post-processing. Furthermore, the commercialization of AutoML in the form of off-the-shelf end-user applications is a likely scenario.

What AutoML is not

The website of the well-respected Machine Learning for Automated Algorithm Design defines Automated Machine Learning as "methods and processes to make Machine Learning available for non-Machine Learning experts."

At face value, this would be an important development.

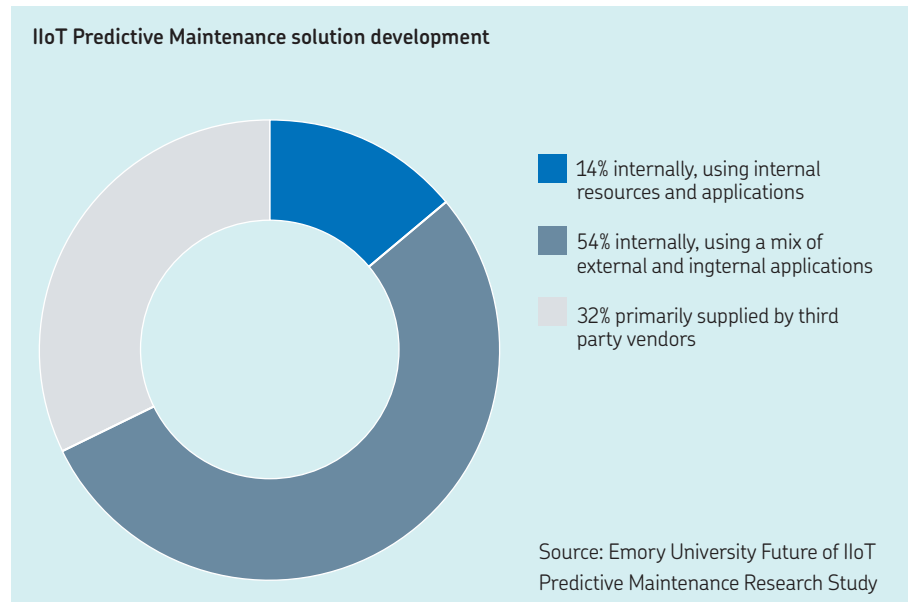
In our opinion, although analysts are forecasting a significant role for "citizen data scientists" who will be able to apply Machine Learning without domain expertise, the current iteration of AutoML is designed to provide tools for non-data scientists. AutoML cannot yet "democratize" the field of Machine Learning.

Is AutoML important for the industrial sector?

According to industry analysts, many senior executives have embraced the Smart Factory concept and IIoT for Predictive Asset Maintenance. As organizations move to implementation, decisions must be made about both the infrastructure required to support the Smart Factory as well as the applications used for Big Data analysis, Predictive Maintenance, etc.

At the core of the decision-making process is whether an industrial plant builds its own technology internally or purchases a solution from third-party vendors.

In the Emory University Future of IIoT Predictive Maintenance Research Study sponsored, Maintenance and Reliability professionals were asked to select the model that will be used to develop IIoT Predictive Maintenance solutions. Only 14% of survey respondents indicated that solutions will be developed internally using only internal resources and applications. Most respondents (54%) believed that solutions will be developed internally but will employ a mix of internal and external applications. Almost one-



third (32%) expected that solutions will be supplied by third-party vendors.

This data reflects the reality that most industrial plants lack the internal competencies to build their own Machine Learning for Predictive Maintenance solutions. AutoML will play an important role in data science in the near future. However, AutoML is not a shortcut that industrial plants can use to bypass the need for Big Data expertise in the Build versus Buy calculation.

Randy Olsen provides the following realistic perspective:

“I don’t see the purpose of AutoML as replacing data scientists, just the same as intelligent code autocompletion tools aren’t intended to replace computer programmers. Rather, to me the purpose of AutoML is to free data scientists from the burden of repetitive and time-consuming tasks ... so they can better spend their time on tasks that are much more difficult to automate.”

The bottom line? Given the limited number of available big data science professionals, third-party applications based on AutoML are the only viable solutions to scale ML-based Predictive Maintenance.

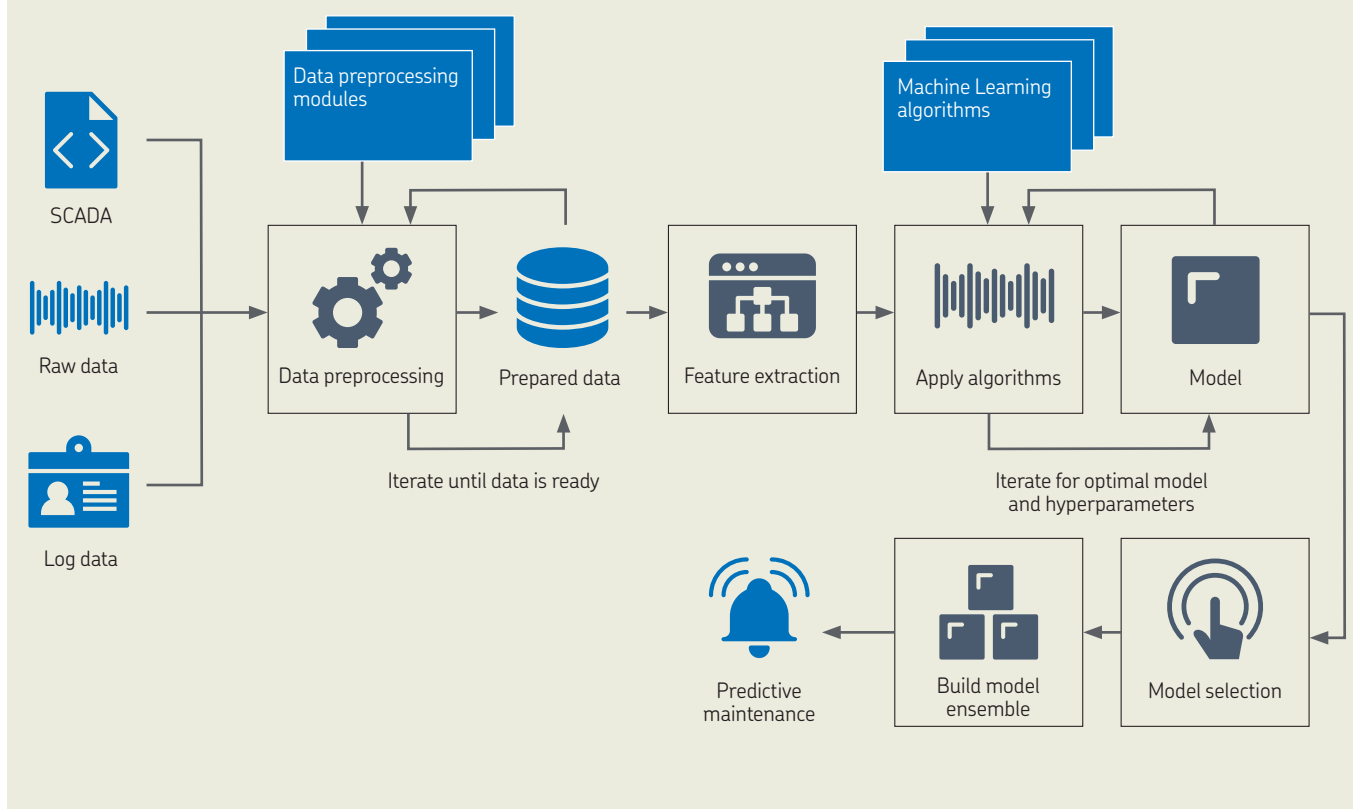
The fallacy of adding computational power

GPU (Graphic Processing Unit) is used to boost computation of Machine Learning. Whereas a CPU is designed for a broad range of tasks, GPU is optimized exclusively for data computations. With cloud-based solutions, on-demand access is provided to a computational power that is distributed amongst a cluster of machines.

Boosting hardware capabilities or accessing cloud-based computational power enables Machine Learning processes such as data pre-processing or hyperparameter optimization. In isolation, this enabler is of limited value without an accelerator such as AutoML that is reviewed in this article.



SKF Enlight AI's automated machine and deep learning engine



Introducing SKF's Enlight AI's AutoML solution

SKF Enlight AI's IIoT Predictive Maintenance solution incorporates key elements of Automated Machine Learning in our methodology. This relates to both the speed of model development and the optimization of model selection.

Enlight AI's AutoML solution contains a full Machine and Deep Learning pipeline responsible for missing values, categorical features, sparse and dense data, and the rescaling/re-treading /normalization of data.

Next, the pipeline applies pre-processing and cleaning algorithms and an ML/DL algorithm.

Enlight AI's AutoML includes dozens of ML algorithms, tens of pre-processing methods, and all their respective hyperparameters, yielding a total of hundreds of hyperparameters.

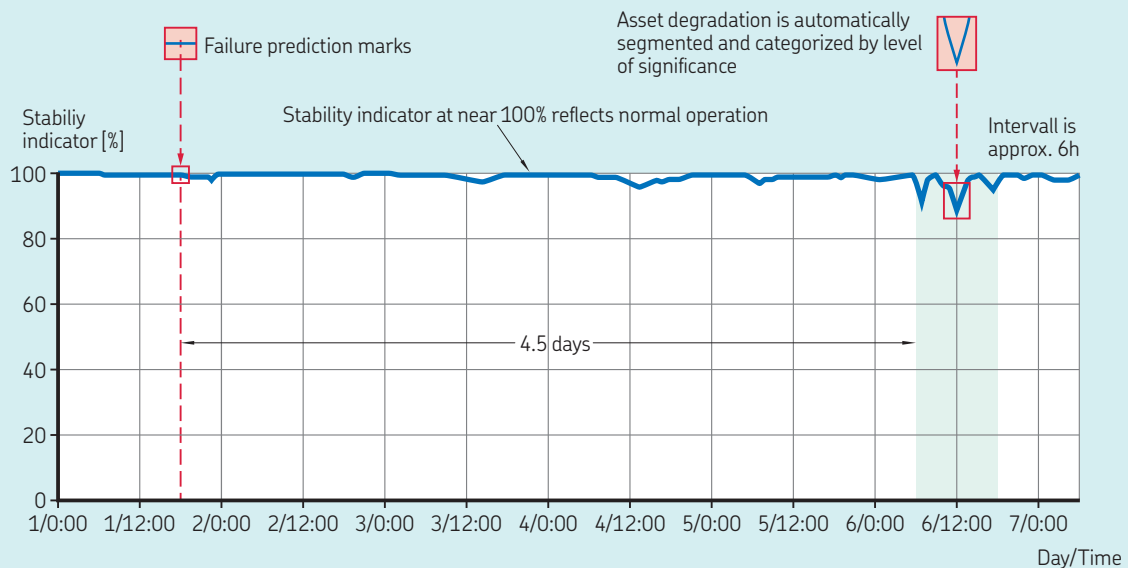
The optimizing performance in AutoML's space of hundreds of hyperparameters can be slow.

SKF Enlight AI jumpstarts this process by using meta-learning to start the job from good hyperparameter settings originating in previous similar datasets (using a similarity function). When a new dataset is added, the algorithm

looks for similar datasets as a starting point, then applies the settings from the previous data set to the new one.

A second improvement is to automatically construct ensembles. Instead of returning a single hyperparameter setting (as standard Bayesian optimization would), we automatically construct ensembles from the models trained during the Bayesian optimization. Specifically, we use model ensemble and stacking to create small, powerful ensembles with increased predictive power and robustness.

SKF Enlight AI's automated Machine Learning solution



This graph depicts the solution's real-time continuous assessment of an asset's health, and the exact moment at which an abnormality is detected in the data and an alert is issued.

Conclusion

Can industrial facilities build AutoML capabilities internally? Only if they already have deep expertise in Machine Learning and can apply the automation of repetitive tasks to speed the company-wide deployment of IIoT PdM.

AutoML benefits the industrial arena when it is part of an IIoT Predictive Maintenance solution. Enlight AI's IIoT Predictive Maintenance incorporates AutoML as a core element of its software.

We expect that the potential of AutoML will gain recognition outside the data science community as SKF Enlight AI continues its mission of applying advancements in Machine Learning to the field of Predictive Maintenance.



skf.com/SKF-Enlight-AI

® SKF is a registered trademark of the SKF Group.

© SKF Group 2020
The contents of this publication are the copyright of the publisher and may not be reproduced (even extracts) unless prior written permission is granted. Every care has been taken to ensure the accuracy of the information contained in this publication but no liability can be accepted for any loss or damage whether direct, indirect or consequential arising out of the use of the information contained herein.

PUB SR/S9 18927 EN · July 2020
Certain image(s) used under license from Shutterstock.com.