

# Automated machine learning (AutoML)

Author, Akash S M

A Data Science Foundation Blog

November 2021

-----  
[www.datascience.foundation](http://www.datascience.foundation)

Copyright 2016 - 2017 Data Science Foundation

In today's world, machine learning is the most popular technology! It is now used in practically every field imaginable, increasing its significance exponentially. But what about individuals who aren't as familiar with Machine Learning? That's where AutoML, or automated machine learning, comes in!

Machine learning, or ML, has been seen and demonstrated time and time again in recent years. However, the application of this huge and sophisticated technology is limited to the small but growing number of data scientists and machine learning enthusiasts and researchers. The notion or concept of Automated Machine Learning was developed to fill this gap and we'll throw some light on the same.

### ***What is AutoML?***

AutoML is a method for automating the entire process of using artificial intelligence (AI) to solve a problem. Data scientists are often in charge of creating machine learning models and all of the associated duties, such as data pre-processing, feature engineering, model selection, hyperparameter optimization, and model post-processing. AutoML frameworks automate these stages (or at least part of them) so that persons without a background in data science can create successful machine learning models.

Automated machine learning (AutoML) is the process of automating the application of machine learning (ML) models to real-world situations. It automates the selection, composition, and parameterization of machine learning models, in particular. Machine learning algorithms that are automated are more user-friendly and often produce faster and more accurate results than hand-coded algorithms.

### ***What is the AutoML process and how does it work?***

Reinforcement learning and a recurrent neural network are used in the logic of AutoML. RNN will first propose a random collection of hyper-values such as nodes per layer, layer count, and so on, and then the model will be built using these parameters.

From processing a raw dataset to deploying a viable machine learning model, AutoML is a platform or open-source framework that makes each stage of the machine learning process easier. Models are created by hand in traditional machine learning, and each step in the process must be handled separately.

---

### ***Data Science Foundation***

## ***What role does automated machine learning play?***

Auto machine learning enables any organisation or corporation to apply machine learning solutions without having to invest additional time and money in locating all of the specialists who can provide a greater return on investment.

Companies in all areas, including health care, can benefit from automated machine learning. Financial markets, fintech, finance, public sector, marketing, retail, sports, manufacturing, and more technologies previously only open to enterprises with massive resources at their disposal to harness machine learning and AI technology.

Automated machine learning assists enterprise users in swiftly adopting machine learning solutions by automating most of the modelling processes required to construct and deploy machine learning models, allowing a company's data scientists to focus on more complex issues.

*Finally, we'll look at the AutoML system's applications, benefits and challenges.*

## ***Which machine learning tasks should be automated?***

While not all aspects of machine learning can be automated, many processes and procedures, particularly in model training, can. These iterative phases are well-suited to automating:

- **Transfer learning and models that have been pre-trained:** Transfer learning in machine learning is adopting models that have already been trained on similar data sets and applying them to your machine learning project. This model is typically used as a starting point and then fine-tuned to meet your specific requirements.
- **Selection of algorithms and hyperparameter optimization:** AutoML programmes can choose the best hyperparameters and algorithms without the need for human interaction.
- **Feature engineering:** AutoML can automate this process of examining input data to develop features that are more compatible with machine learning algorithms.
- **Data pre-processing:** It entails procedures such as data cleaning, data integration, data transformation, and data reduction to improve data quality and convert unstructured, raw data into a structured format.

## ***What are the Benefits of Using AutoML Tools?***

When we look at the present procedure for creating a machine learning model, we see that it

often requires highly experienced technical experts, a lengthy development process, a large sum of money, and a large number of iterations. All four of these causes are driving the need for AutoML:

### **Data that is well-organized**

AutoML is best suited for machine learning applications that employ structured data, which covers the vast majority of tasks. When features are sorted into rows and columns, they are automatically structured for consumption by AutoML tools. It doesn't matter if any data in the columns is missing because AutoML tools handle imputation. To engineer features for ML algorithms, AutoML tools will encode categorical variables and normalise numerical ones.

### **It helps you save time**

No one is born knowing the best algorithm and hyperparameters for addressing a problem. Instead, we manually test models, modify hyperparameters, and evaluate models to find the optimal solution for a given problem. This manual labour is abstracted away by AutoML. AutoML assists you in transferring your data to the training algorithm and searching for the appropriate neural network design for your problem automatically, saving you time. You can receive results in 15 minutes instead of hours with AutoML.

### **Reduced Errors When Using Machine Learning Algorithms**

As companies grow, market trends change, and the amount of data available grows. AutoML improves models by minimising the likelihood of inaccuracies caused by bias or human mistakes. Businesses may innovate with confidence, achieve more accuracy, provide business benefits, and increase ROI on ML projects with this edge.

### **Challenges of AutoML**

AutoML systems can already generate predictive models quickly and with near-optimal results. However, their coverage is still limited, and their full potential has yet to be realised.

*In this part, we'll go through three drawbacks of existing AutoML systems:*

### **Conformance to flexible specifications**

While most AutoML solutions stress performance, in the real world, performance is only one

component of machine learning projects. Again, you can't simply turn your back on the human intelligence buried in machine learning models. As a result, the computing and storage specification sheet cannot be compromised by the businesses.

## **The 80/20 rule**

Automated machine learning automates the majority of the work, which you could accomplish yourself in many circumstances. Regardless of whether or not automated machine learning is used, the 20% will take a significant amount of effort. A similar concept might be used for businesses. The majority of data science work nowadays is focused on financial services, insurance, and other similar industries. In any event, if one's industry is outside of this, one may have fewer prebuilt components.

## **Explainability**

Although automated robots can identify solutions, they may not always be what the user desires. A user may desire a model that is easy to understand. Explainability, experts believe, is fraught with uncertainty because everyone's comprehension differs, and it has a strong link to personal judgement. Making the model explainable is considerably more difficult.

## **Top Well-known AutoML Tools**

*Let's look at some of the most popular tools for automating parts of or the complete machine learning pipeline. This isn't a full list, but we tried to include the ones that are commonly used:*

### **H2OAutoML**

It's a scalable distributed in-memory machine learning platform. It's another open-source application with a plethora of automated machine learning modules. It can automate the entire machine learning workflow, including model tuning and automatic training, within a set time constraint. H2O.ai can also fully automate some of the most time-consuming and difficult data science operations, such as model assembly, feature engineering, model tuning, and model deployment.

### **SMAC**

SMAC (sequential model-based algorithm configuration) is a powerful tool for adjusting algorithm parameters. SMAC is a powerful tool for hyperparameter optimization in machine

learning algorithms, scaling better than other techniques to large dimensions and discrete input dimensions.

### **Auto-Pilot by Amazon Sagemaker**

The Sagemaker Auto-Pilot is a component of the AWS Machine Learning framework Sagemaker. It provides a platform and tools to support an ML project across its entire life cycle. For the following three challenges, Amazon Sagemaker Auto-Pilot recommends utilising AutoML: classification, regression, and time-series prediction.

### **Microsoft Azure AutoML**

For customers who are unfamiliar with coding, Azure AutoML provides a clear model selection procedure. It is a cloud-based service that allows users to create and manage machine learning solutions. Azure is a platform that can be learned without prior programming experience.

### ***Will AutoML take the place of data scientists?***

In a word, no. While AutoML tools are excellent, they only help data scientists with one aspect of their job: model optimization. AutoML can automate the model-optimization technique since it is reproducible across a wide range of use cases. To find use cases, obtain sufficient data, design applications, pick metrics, and so on, data-science skills are still required. AutoML technologies can also be used by data scientists to automate some of the tedious aspects of their jobs, allowing them to focus on more complicated challenges.

### ***EndNote: A Leap into the Future!***

To summarise, we believe AutoML is an intriguing AI solution that will lower the AI entry barrier in many scenarios. As a result, both prototyping and small-scale AI solutions can be deployed at a low cost. Of course, there are disadvantages. It's important to think about the inflexibility and knowledge gap that come with not implementing machine learning in-house.

Nonetheless, AutoML is a step toward automating the entire process of applying AI to verifiable problems. AutoML, we believe, will have a stronger foothold in the AI industry in the next few years.

## About the Data Science Foundation

The Data Science Foundation is a professional body representing the interests of the Data Science Industry. Its membership consists of suppliers who offer a range of big data analytical and technical services and companies and individuals with an interest in the commercial advantages that can be gained from big data. The organisation aims to raise the profile of this developing industry, to educate people about the benefits of knowledge based decision making and to encourage firms to start using big data techniques.

## Contact Data Science Foundation

Email: [admin@datascience.foundation](mailto:admin@datascience.foundation)

Telephone: 0161 926 3641

Atlantic Business Centre

Atlantic Street

Altrincham

WA14 5NQ

web: [www.datascience.foundation](http://www.datascience.foundation)

---

### ***Data Science Foundation***

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ

Tel: 0161 926 3670 Email: [admin@datascience.foundation](mailto:admin@datascience.foundation) Web: [www.datascience.foundation](http://www.datascience.foundation)

Registered in England and Wales 4th June 2015, Registered Number 9624670