

Machine learning with AutoGluon, an open source AutoML library

by Shashank Prasanna | on 31 MAR 2020 | in [Amazon EC2](#), [Amazon Machine Learning](#), [Open Source](#), [Technical How-To](#) | [Permalink](#) | 

[Comments](#) | [Share](#)

If you work in data science, you might think that the hardest thing about machine learning is not knowing when you'll be done. You start with a problem, a dataset, and an idea about how to solve it, but you never know whether your approach is going to work until later, after you've wasted time. Part of what makes the machine learning process difficult is that there are a lot of best practices that experienced practitioners know to use. If you're just getting started in data science, you may spend a significant amount of time on an approach you thought was right, which an expert practitioner would have told you is a dead end.

What if you could codify these best practices into one simple and easy-to-use software package that any developer could use? A library that can automatically prepare your dataset, try different machine learning approaches, and combine their results to deliver high-quality models—and all of that with a few lines of code?

This is the idea behind automated machine learning (AutoML), and the thinking that went into designing [AutoGluon](#) AutoML library that [Amazon Web Services \(AWS\)](#) open-sourced at [re:invent 2019](#). Using AutoGluon, you can train state-of-the-art machine learning models for image classification, object detection, text classification, and tabular data prediction with little to no prior experience in machine learning. You can run AutoGluon anywhere—from your laptop or workstation, to a powerful [Amazon Elastic Compute Cloud \(Amazon EC2\)](#)—instance to take advantage of multiple cores and get results faster.

The AutoGluon team at AWS has [released a paper](#) detailing the inner-workings of AutoGluon-Tabular, an open source AutoGluon capability that allows you to train machine learning models on tabular datasets from sources such as spreadsheets and database tables.

In the first half of this article, I will introduce AutoGluon-Tabular and summarize key innovations described in the paper and the magic that happens behind the scenes when you use AutoGluon-Tabular. In this second half of the article, I will walk through an end-to-end code example showing how you can use AutoGluon-Tabular to get top 1% scores in a data science competition with a few lines of code—no machine learning experience required.