

AutoML Is Not Enough

Whitepaper

20
20

Big Squid AutoML Overview:

AutoML is an excellent tool to enable data scientists and analysts alike to more effectively bring value to their organizations, because it so heavily augments their intuition and supports them in speedy problem solving—but is it enough?

AutoML—**automating the process of machine learning to solve real-world problems**—has gained popularity because it makes data science easier, faster, more affordable, and less cumbersome. Considering data scientists are in high demand but low supply, they are often overburdened by their workloads. AutoML is an excellent tool to enable data scientists to more effectively bring value to their organizations, because it so heavily **augments their intuition** and **supports them in speedy problem solving**—but is it enough? **No! AutoML on its own is not enough** to support machine learning in the modern business landscape, one which moves and evolves at an ever-quickenning pace while drowning in veritable oceans of data. What we need is a paradigm shift.

But, Big Squid! You guys do AutoML.

Yep, okay, that's completely true, but we also do more than that. We are solving the obstacles in data science and machine learning from the ground up. We believe machine learning should be understandable (e.g. **explainable and interpretable**), that it should be available and usable by everyone—even people who don't have a PhD from MIT—and that ML and humans should work closely together to augment one another's abilities. In order for all of that to work, we have to start with the very foundations of machine learning in the business world—the data science workflow and the people who work within it.

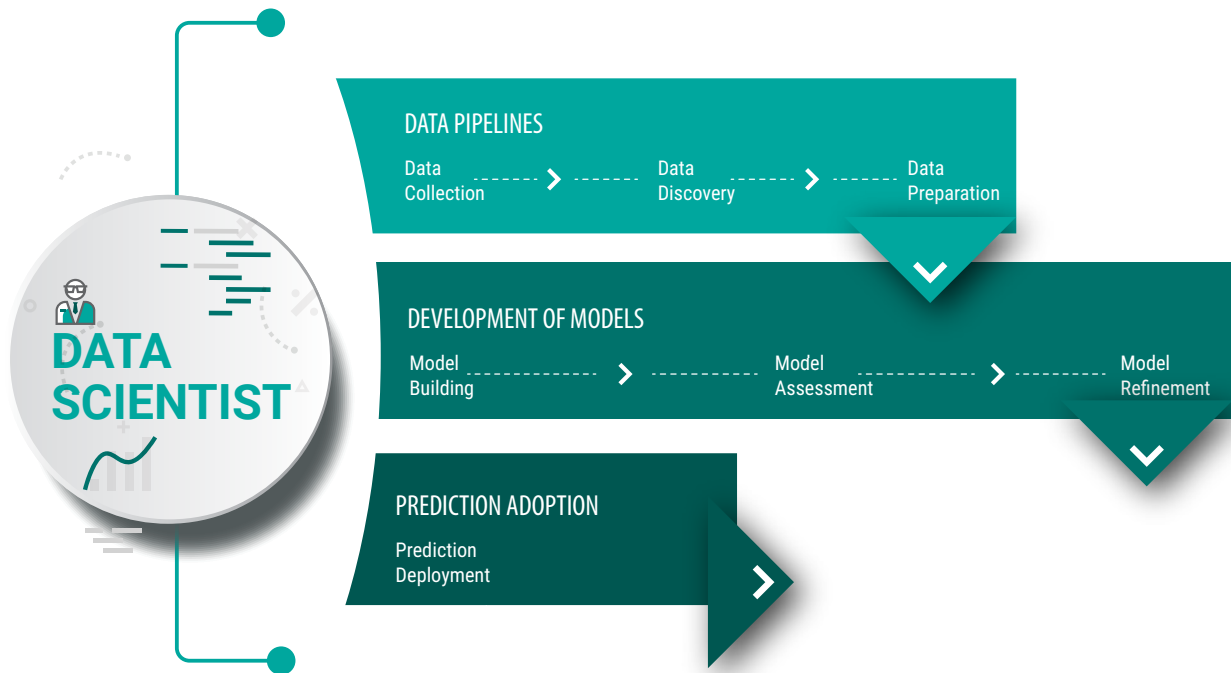
The data science workflow is broken.

There, we said it. Until the data science workflow itself supports machine learning in an effective way, no amount of AutoML will support the demands of today's data-driven world. In the traditional data science workflow, AutoML is just a band-aid on a bullet wound. However, we propose a solution. We can fix the data science workflow if we simply look at it from a different angle and take action.



THE TRADITIONAL DATA SCIENCE WORKFLOW AND ITS FLAWS

When most people think about the data science workflow, it is from the perspective of the data scientist. It usually looks something like this:

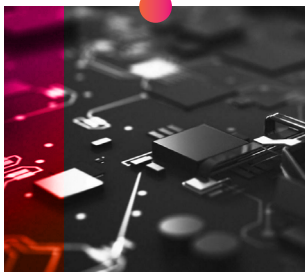


However, in the real world, the tasks required to build an effective, valuable, scalable machine learning model do not begin and end with the data scientist, so this structure is flawed from the start. It is overly focused on building a model and getting it into production, even though productionalization on its own in no way ensures success.

The traditional data science workflow model is flawed in three essential ways:

1. It does not ensure adoption, because it ignores upfront business objectives.
2. It does not ensure business value, because it does not focus on the action steps needed to derive value.
3. It is not scalable, in part because it is not inclusive of all stakeholders.

We will cover each of these in greater depth.



ADOPTION

A meager 13% of machine learning models are put into production. For the amount of money being invested in machine learning projects, this is a big problem. Why is there so little adoption?

This usually comes down to failing to engage business decision-makers. This can range from analysts, data engineers, stakeholders, and executives, and the problem is usually most apparent at the very early and very late stages of development—two points when cohesion of goals are most relevant.

In the early stages, it is vital to frame—very specifically—the business goal with which the machine learning model is intended to align. Yet, the step is often overlooked, instead replaced by a mountain of data and people sitting around going, “Hmm, what can we figure out using this?” And so begins the cycle of excess work on a project that may never even be adopted because its... useless is perhaps a strong word, but irrelevant at best.

Ideally, data preparation comes only after the key business objective is clearly defined along with specific information about what actions to take in order to generate value.

At a later stage, once the model has already been created, failure to adopt also hinges heavily on communication between business decision-makers. Executive sponsors, analysts, stakeholders, and data scientists must all be able to discuss and connect with one another in accessible language in order to make the model actually usable—i.e. worth adopting. If everybody within this circle doesn’t understand the how, the why, and the “what next?”, then the model is already dead in the water, sunk costs and all.

Understanding the best ways to present outputs from an ML model to “non-data” team members is critical. Especially for business leaders who are used to making decisions based on data through business intelligence dashboards, such as those within widely-used ERP software, it might be best to present

predictions in ways they are already familiar with or can easily adapt to, rather than hand them a funky looking science project and hope everyone is on the same page. As with most things in the business world, framing and presentation are vital to impact and adoption.

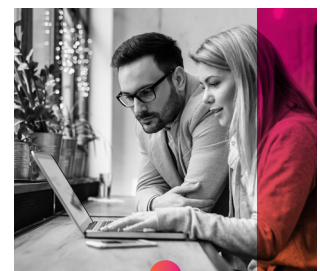
By congregating multiple business roles in the ML and data science process, businesses are able to generate personal investment in the success of the project. Considering all machine learning models require multiple iterations to achieve ideal functionality, ensuring everyone understands the value, methods, and actions associated with a machine learning model helps ensure its ongoing adoption, success, and, ultimately, its ROI. Understanding and iteration also help teams trust the model's outputs, a requirement for taking meaningful action and driving business value.

BUSINESS VALUE

The failure of the traditional data science workflow to provide real business value hinges on many of the same things that serve to undermine adoption, only much more compounded. If few ML models ever make it to production, then, of course, even fewer truly succeed and drive value for the business—but why? What stops businesses from extracting value from machine learning initiatives?

In order to derive business value from machine learning, you must know:

- **Why?** What key business objective is the machine learning model meant to serve? How is meant to do that?
- **Then what?** What within the machine learning model triggers action? What actions can be taken? Basically.... What will you do with the insights?





In the traditional data science workflow, it is easy to get bogged down in creating and perfecting a model. While this might be purposeful as a learning exercise or advancement opportunity in a research setting, it basically equates to expensive tinkering in the business setting. (Ouch.) Collecting and perfecting can not supersede the need for action.

Business users of ML must have an action plan in place that tells them how to assimilate and process insights from the machine learning model, when to take action based on these predictions—what is the trigger?—, what actions could be relevant to take, and what key business objectives these actions impact, and why.



A **real-world example** of this is routing a customer to the support team when their predicted risk of churn passes a certain threshold. It could also mean kicking off a tailored email campaign when an inbound lead is predicted to fit a specific customer profile. Why route a customer to support when they pass that threshold? Because “what-if” modeling indicated when support reaches out to that specific type of customer, their risk of churn dissipates substantially, increasing customer retention and income for the business.

AutoML can certainly help with value extraction, because it increases efficiency both in terms of time and money—but if and only if it is utilized in a way that lays the foundation for value extraction and within a workflow that plays into that objective. When all the pieces come together in this way, stakeholders and decision-makers rest easier with greater clarity and confidence in their decisions, rather than toss and turn trying to suss out benefit from “perfect” but confusing and expensive ML models.

SCALABILITY

Data scientists are unicorns with difficult-to-find skills packaged in a single resource and tied up with a bow. Everybody wants one, but when they get one—well, businesses tend to use and abuse their data scientists with hefty workloads that no one man or woman should have to shoulder alone. This inevitably leads to **burnout**, loss of motivation and creativity, and a vastly lowered potential for business value.

With so many business objectives that machine learning can benefit, it's no wonder why most data scientists have a huge backlog with senior-executive backed projects often cutting to the front of the line. With so many breaks in focus and priority, while the data scientists juggle a job fit for a team, it's obvious why the average machine learning project takes **6–9 months** to get to production and **costs \$250k**.

Data scientists are an obvious bottleneck that prevents scalability. Nothing so beneficial and far-reaching as machine learning should be constrained to such a scarce and precious resource. Rather, through **data democratization** and better utilization of employees which we will cover in greater depth below, data scientists can be preserved for only the most complex and pressing tasks, backed by a full team of support. **Only when we remove the isolation and siloed nature of data science and ML can it be scalable and reach full value potential. Collaboration is critical.**



Only when we remove the isolation and siloed nature of data science and ML can it be scalable and reach full value potential.

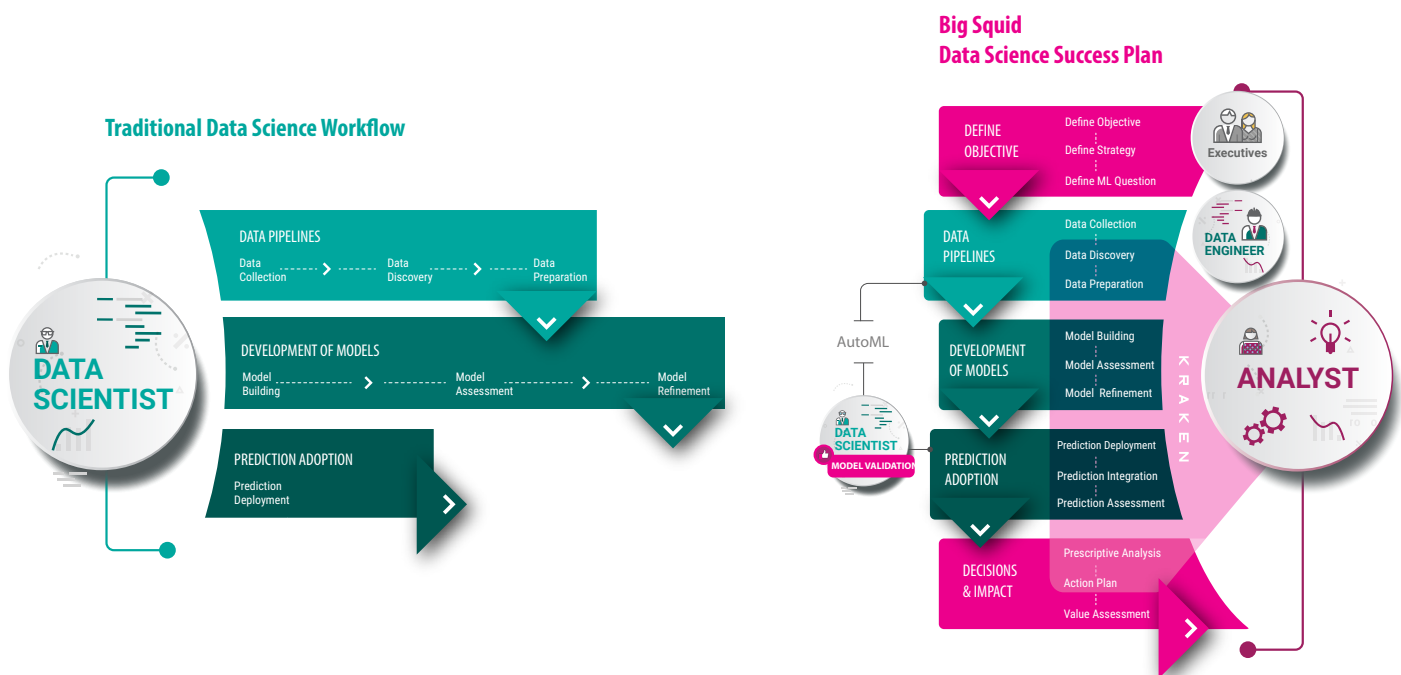
Collaboration is critical.

DATA SCIENCE WORKFLOW V2.0 - NEW & IMPROVED

There is a better approach to the data science workflow filled with collaboration, action, and real, serious value. In this new model, traditional data science teams come together with key stakeholders from the outset to clearly define and understand business objectives. With this information, leaders can determine a clear course of action towards what questions can be answered using machine learning in order to drive the most value for the organization, how to know when to take action, and a follow-up plan for actually doing it.

In this model, everyone understands what they are working toward and why, and it also establishes a clear baseline for communication which helps to ensure interpretability and explainability at each stage of the process. This new model also engenders a sense of **solidarity** and teamwork that is key to positive business culture and utmost job satisfaction and productivity and that will ultimately keep teams cohesive and engaged at the point of productionalization and iteration. Similarly, when everyone is on board and understanding of the place of machine learning, it assists with confidence in ML predictions and confidence in taking action on those predictions.

Here's what it looks like:



UNDERUTILIZED EMPLOYEE POTENTIAL AND A HOME FOR AUTOML

We've already determined that data scientists are a scarce and precious resource to be handled with the type of care one would give to The One Ring To Rule Them All, so we definitely don't want them ugly-crying in the bathroom under the sheer stress of their workload as they tilt towards a nervous breakdown. In order to prevent that, we have to provide data scientists the support they need to flourish in the jobs they do best. Luckily, that support may already be in the cubicle next door—all it takes is a little enablement, a little freedom, and a healthy dose of AutoML.

In the first step of the new data science workflow, we need **key stakeholders** and decision-makers to help frame important business objectives. This lays the framework for valuable problem-solving. Stakeholders can assist with defining the objective, strategy, and key question to answer with ML, as well as how and when to take action once we have answers.



Then, data scientists can happily clickety-clack away at their desks to figure out what data to gather and where and how to get it. Alongside gaining insight into the data pipeline, data scientists can also help with model validation—a critical step in the success of a machine learning initiative, perhaps the most critical step at which to utilize the mythical and highly-sought data scientist unicorn. This is their natural habitat, and you will often find they thrive here.

*But—this is important—they don't have to do **everything else** and they don't even have to do those steps alone.* They should certainly play a role in understanding business objectives, and through **data democratization**, analysts can readily assist with data exploration, collection, and structuring. With the addition of AutoML like Big Squid's Kraken, a support team of analysts can shoulder a lot of the load, freeing up the bottleneck, and allowing for faster, easier, more effective machine learning across the board. Analysts, and sometimes other data science roles, can help with model building, assessment, and refinement. They can also help and the prediction and adoption phase with deployment, integration, and assessment. Then they can help even more with prescriptive analysis, building a deeper action plan based on the model's insights, and then again with value assessment.



**Achievement unlocked:
AI simplified.**

**Smarter decisions,
faster.**

At Big Squid, we can help you verify your readiness for machine learning then help put your business on a fast track to success toward your machine learning initiatives with guidance down the path of the new data science workflow. Our team can help you find out which questions ML can answer based on key business objectives, then supercharge your efforts by providing the AutoML software to support your journey (and your data scientists).

AutoML like Kraken empowers data analysts across the entire data science workflow to make notable impacts and take on a much greater role in the machine learning process both by simplifying that process and augmenting the skills of the analysts. Analysts basically become “citizen” data scientists, freeing up the actual data scientists to do only the most high-level, targeted work—streamlining the whole process from the ground up. With a good data science workflow accompanied by AutoML, everyone goes up a level. **Achievement unlocked: AI simplified. Smarter decisions, faster.**



www.bigsquid.com