

OLD WINE IN SMART BOTTLES: THE MOTLEY

by [Patrick Michl](#)

25. March, 2019

Since almost 10 years the rate of data science publications has been growing enormously! For scientists and developers, it is therefore becoming more and more difficult to keep track of suitable current approaches.

The accelerated publication rate, is not only caused by an increased demand, but also due to the very nature of this area: The more you already know about your data domain (wise people call it “belief”), the better your estimations for the observed sample can be. This simple Bayesian wisdom has great implications for data science: The publications are not only getting more in numbers, but also in the numbers of partially overlapping domains!

So it should be quite clear, what the issue is. But what are the current tools, to address it? Corresponding to their individual data domains, the research papers in data science are distributed among different platforms. A typical example, that I would like to pick is [arXiv](#) (pronounced archive). Since the early days of the web, this ‘Grande Dame’ essentially serves as a plain repository for PDF pre-prints. And since data science papers usually deal with algorithms, these PDFs are quite often endowed with pseudo code, which basically allows their implementation in any programming language of choice.

It is quite obvious, that this organically grown [Paper Bottleneck](#) has substantial

drawbacks: Due to the limited space, the provided pseudo code often loses valuable details over the original algorithm. But also if the original algorithm is provided online, it can be grueling to properly identify it’s scope and adapt it to the underlying prerequisites - only to decide about it’s suitability!

What is Motley?

In a nutshell: The development- and exploration process in data science is currently heavily impaired by the detour experienced by publications in paper form. While it became easier and easier to publish, it became harder and harder to get an overview. This is why we want to provide a better solution!

Motley is a smart algorithm repository server, that enforces unified data interfaces for different algorithm categories. This allows Motley not only to automatically evaluate and compare the hosted algorithms with respect to given metrics but thereupon also to determine, which algorithm of a given category and data domain is the currently best fitting (CBF) algorithm with respect to some metric. An example for such a metric would be the average prediction accuracy within a fixed set of gold standard samples of the respective domain of application (e.g. latin handwriting samples, spoken word

samples, TCGA gene expression data, etc.).

According to our [convictions](#), Motley is [free and open-source](#), based on the

[Python](#) programming language and actively developed as part of our [Liquid Coding](#) framework. For more information please visit [GitHub](#).