* A more rigorous way to view benchmarking is as an approach to identify the respective strengths weaknesses of a given methodology in contrast with others Caruana and NiculescuMizil (2006) PMLB
* Benchmarking could be thought of simply as a sanity check to confirm that a new method successfully runs as expected and can reliably find simple patterns that existing methods are known to identify Hastie et al. (2009) PMLB
* The term benchmarking is used in machine learning (ML) to refer to the evaluation and comparison of ML methods regarding their ability to learn patterns in ‘benchmark’ datasets that have been applied as ‘standards’. PMLB
* It allows us, as a community, to track progress over time, to identify still challenging issues, and to learn which algorithms are most appropriate for specific applications. OpenML 2017
* However, we currently lack standardized, easily-accessible benchmark suites of datasets that are curated to reflect important problem domains, practical to use, and that support a rigorous analysis of performance results. OpenML2017