Building and Benchmarking AutoML Systems

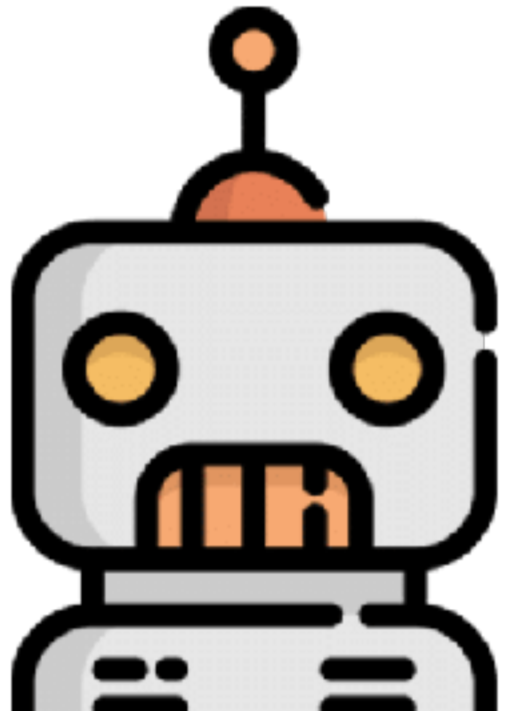
UseR! Toulouse
July 2019

Erin LeDell Ph.D.
@ledell

Agenda

- Automatic Machine Learning (AutoML)
- Machine Learning Benchmarking
- Benchmarking in AutoML development
- Benchmark of OSS AutoML Systems

Slides ⬇️ https://tinyurl.com/user19-amlbench



Automatic Machine Learning (AutoML)

Goals & Features of AutoML

- 🏆 Train the best model in the least amount of time.
- 📉 Reduce the human effort & expertise required in machine learning.
- 📈 Improve the performance of machine learning models.
- 🔄 Increase reproducibility & establish a baseline for scientific research or applications.

Aspects of Automatic Machine Learning

Data Prep

Model Generation

Ensembles

Aspects of Automatic Machine Learning

Data Preprocessing

- Imputation, one-hot encoding, standardization
- Feature selection and/or feature extraction (e.g. PCA)
- Count/Label/Target encoding of categorical features

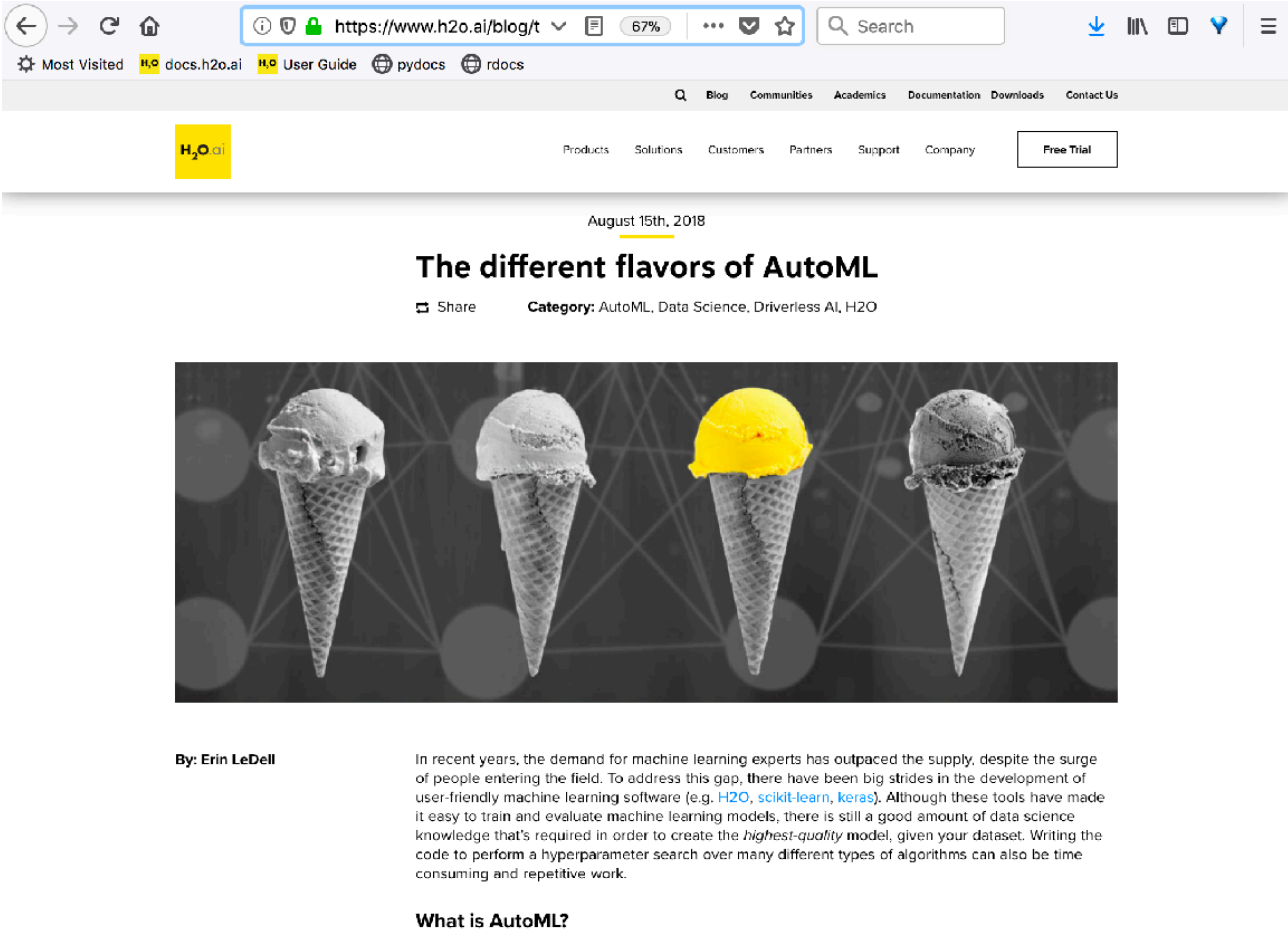
Model Generation

- Cartesian grid search or random grid search
- Bayesian Hyperparameter Optimization
- Individual models can be tuned using a validation set

Ensembles

- Ensembles often out-perform individual models
- Stacking / Super Learning (Wolpert, Breiman)
- Ensemble Selection (Caruana)

Different Flavors of AutoML



August 15th, 2018

The different flavors of AutoML

Category: AutoML, Data Science, Driverless AI, H2O

By: Erin LeDell

In recent years, the demand for machine learning experts has outpaced the supply, despite the surge of people entering the field. To address this gap, there have been big strides in the development of user-friendly machine learning software (e.g. H2O, scikit-learn, keras). Although these tools have made it easy to train and evaluate machine learning models, there is still a good amount of data science knowledge that's required in order to create the *highest-quality* model, given your dataset. Writing the code to perform a hyperparameter search over many different types of algorithms can also be time consuming and repetitive work.

What is AutoML?

https://tinyurl.com/flavors-of-automl

Machine Learning Benchmarking

ML Benchmarking

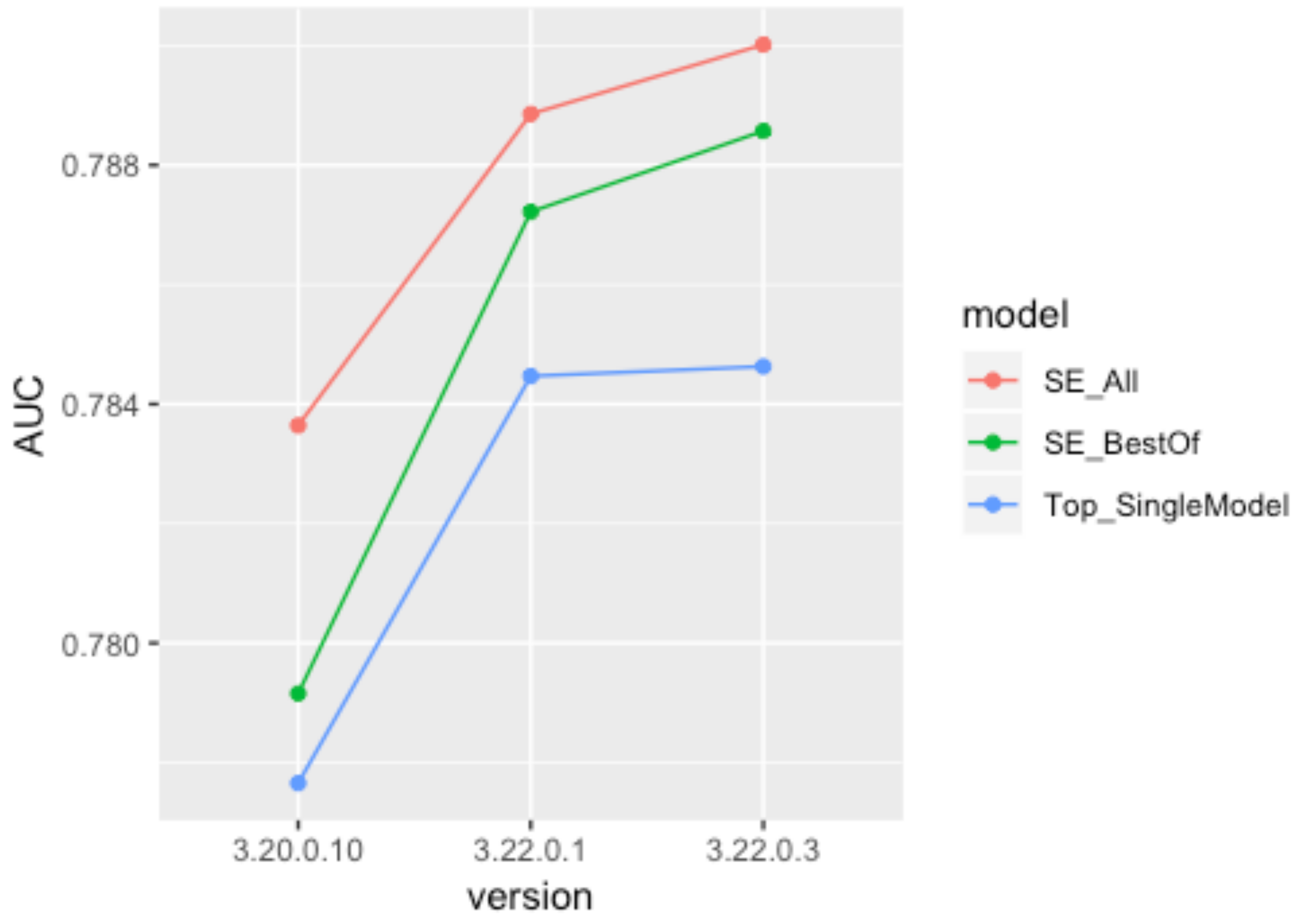
- 📊 Compare model & runtime performance of machine learning tools
- 🔍 Provide accurate information for users to discriminate between tools
- 🔄 Best to run on fixed & publicly available hardware such as Amazon EC2
- 🇨🇭 Best done by a third-party and not an author

ML Benchmarking Mistakes

- Not enough datasets, not enough diversity among the datasets and datasets are too small ❌
- Tools benchmarked incorrectly or unfairly:
  - Package authors are experts at using their own tool but make mistakes using others ❌
  - Inappropriate metrics used ❌
  - Tuning some algorithms more than others ❌
  - Insufficient memory or CPUs ❌
  - Over-generalization of results ❌

Benchmarking for AutoML development

Benchmarking for AutoML



Changes made to the H2O AutoML algorithm and the effect on performance:

- 3.20.0.10 – Baseline
- 3.22.0.1 – Add XGBoost
- 3.22.0.3 – Modify validation strategy

Why is benchmarking so important for AutoML development?

- There is no "reference algorithm" in AutoML so we are creating new methods from scratch.
- It's easy to overfit your tool to familiar datasets.
- Every time you make a change to the algorithm, you should justify the change via benchmarks.

AutoML Benchmark

AutoML Benchmark

OpenML

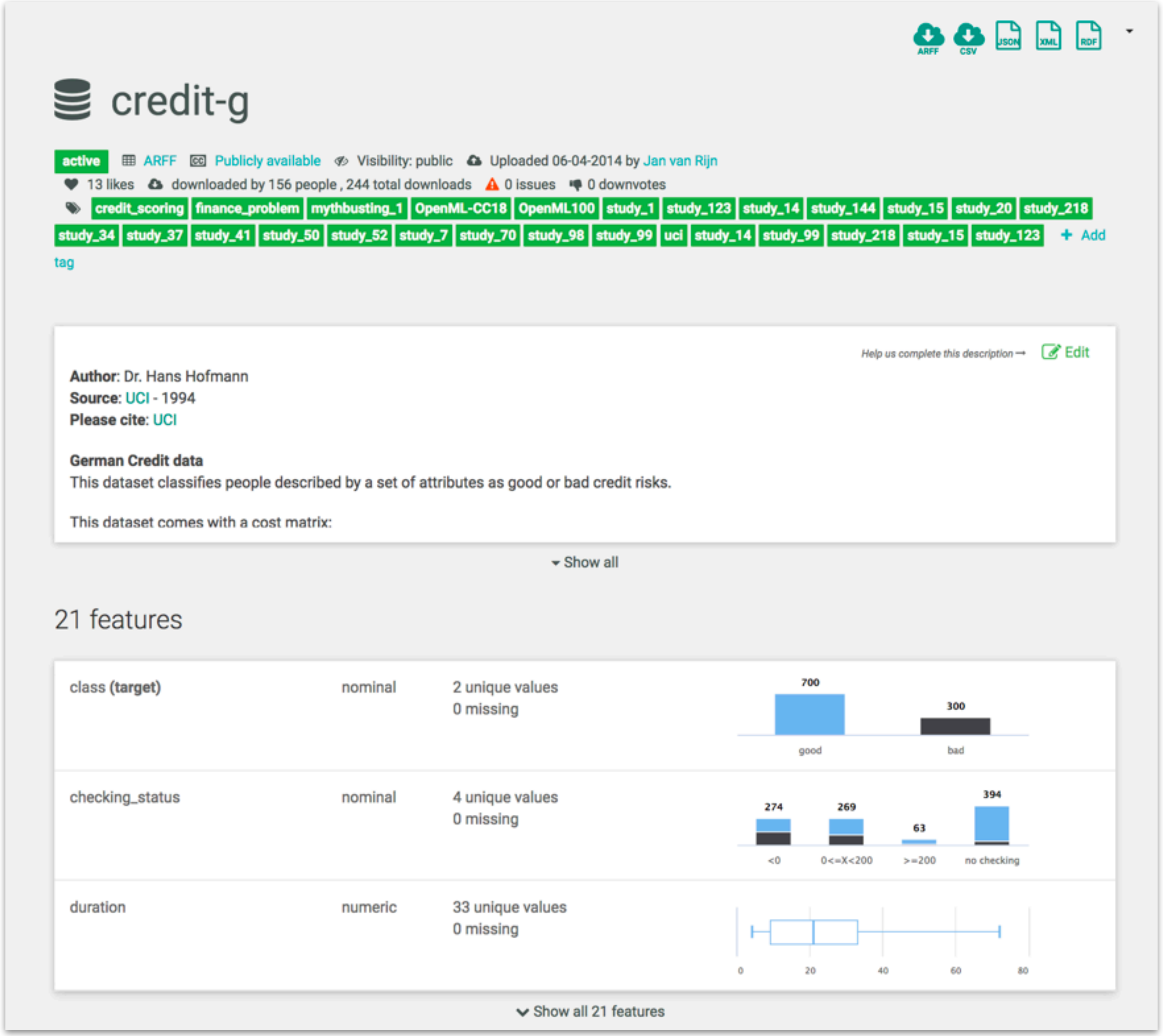
Collaboration between AutoML researchers and OpenML.org to develop a system for high quality benchmarks of the popular open source AutoML systems.

https://github.com/openml/automlbenchmark

OpenML

openml.org

- Platform for reproducible ML experiments
- Unique IDs for datasets & ML tasks
- OpenML data is used in many ML benchmarks



credit-g

active ARFF Publicly available Visibility: public Uploaded 06-04-2014 by Jan van Rijn
13 likes downloaded by 156 people , 244 total downloads 0 issues 0 downvotes
credit_scoring finance_problem mythbusting_1 OpenML-CC18 OpenML100 study_1 study_123 study_14 study_144 study_15 study_20 study_218 study_34 study_37 study_41 study_50 study_52 study_7 study_70 study_98 study_99 uci study_14 study_99 study_218 study_15 study_123 + Add tag

Help us complete this description → Edit

Author: Dr. Hans Hofmann
Source: UCI - 1994
Please cite: UCI

German Credit data
This dataset classifies people described by a set of attributes as good or bad credit risks.

This dataset comes with a cost matrix:

Show all

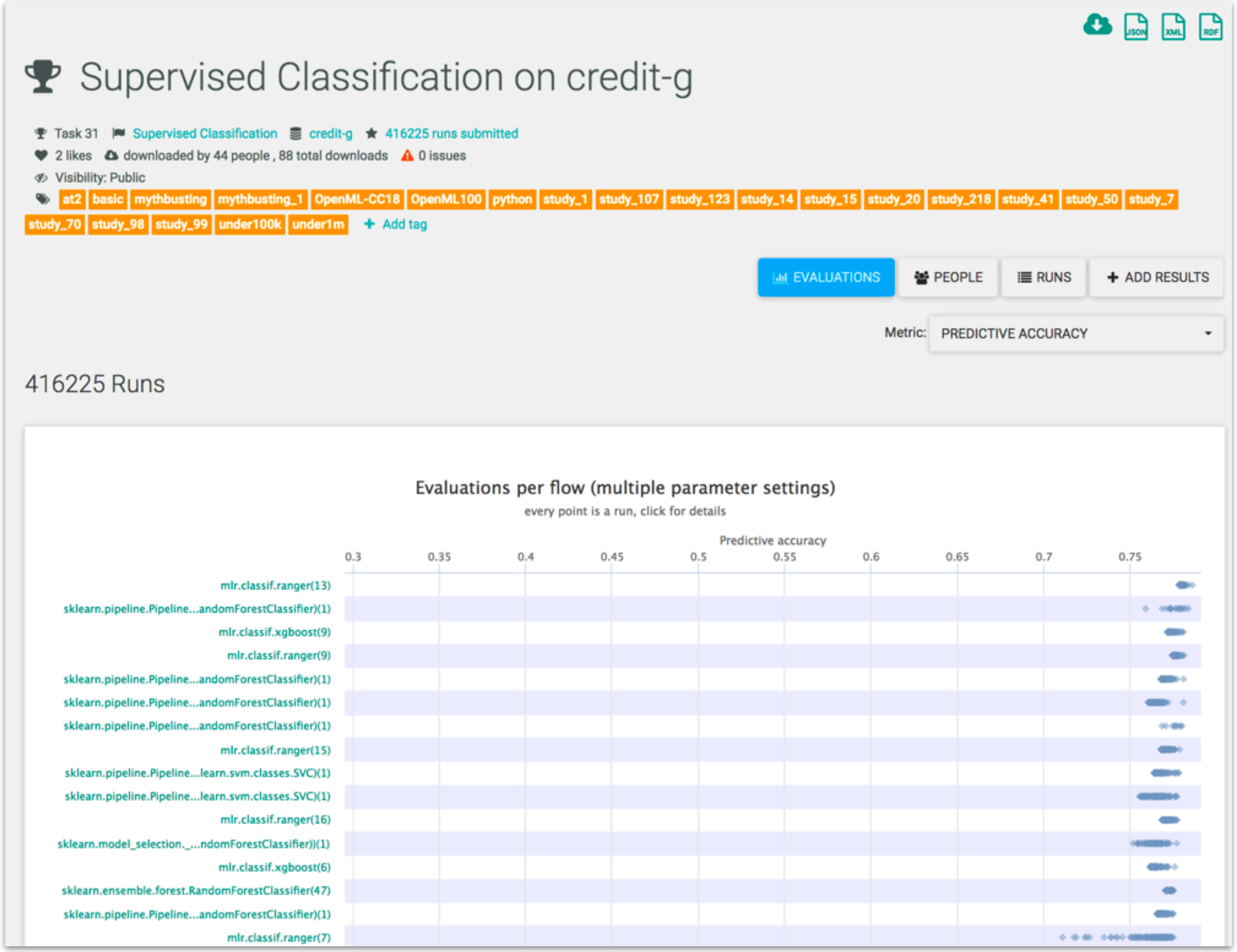
21 features

class **(target)**	nominal	2 unique values 0 missing
checking_status	nominal	4 unique values 0 missing
duration	numeric	33 unique values 0 missing

Show all 21 features

https://www.openml.org/d/31

OpenML



Supervised Classification on credit-g

Task 31 Supervised Classification credit-g 416225 runs submitted

2 likes downloaded by 44 people , 88 total downloads 0 issues

Visibility: Public

at2 basic mythbusting mythbusting_1 OpenML-CC18 OpenML100 python study_1 study_107 study_123 study_14 study_15 study_20 study_218 study_41 study_50 study_7 study_70 study_98 study_99 under100k under1m

Metric: PREDICTIVE ACCURACY

416225 Runs

Evaluations per flow (multiple parameter settings)

every point is a run, click for details

Predictive accuracy
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
mlr.classif.ranger(13)
sklearn.pipeline.Pipeline...andomForestClassifier)(1)
mlr.classif.xgboost(9)
mlr.classif.ranger(9)
sklearn.pipeline.Pipeline...andomForestClassifier)(1)
sklearn.pipeline.Pipeline...andomForestClassifier)(1)
sklearn.pipeline.Pipeline...andomForestClassifier)(1)
mlr.classif.ranger(15)
sklearn.pipeline.Pipeline...learn.svm.classes.SVC)(1)
sklearn.pipeline.Pipeline...learn.svm.classes.SVC)(1)
mlr.classif.ranger(16)
sklearn.model_selection._...ndomForestClassifier))(1)
mlr.classif.xgboost(6)
sklearn.ensemble.forest.RandomForestClassifier(47)
sklearn.pipeline.Pipeline...andomForestClassifier)(1)
mlr.classif.ranger(7)

https://www.openml.org/t/31

OpenML tasks are uniquely defined by dataset & response column, along with evaluation method (e.g. 10-fold CV).

AutoML Benchmark

- 🗄️ Defined a diverse collection of datasets
- 🐳 Open source Dockerized framework for executing benchmarks locally or on Amazon EC2
- ➕ Extensible architecture (easy to add new tools)
- 📊 Results available on the web
- 🔄 Can re-run benchmarks on new tool versions & will expand to more tools, datasets & use cases

AutoML Software

What qualifies as “AutoML” software?

- 👉 Point to a dataset & response column (no other required hyperparameters).
- 🏆 Returns the best model and optionally a list of all models trained.
- ⏱️ Time or resource budget.

Example: H2O AutoML in R

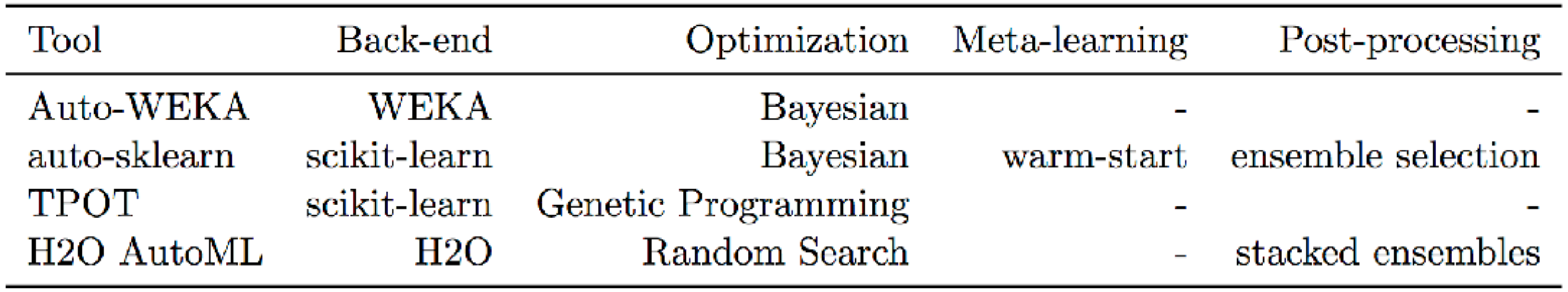
Example

```r
library(h2o)

h2o.init()

train <- h2o.importFile("train.csv")


aml <- h2o.automl(y = "response_colname",
                  training_frame = train,
                  max_runtime_secs = 600)

lb <- aml@leaderboard
```

AutoML Software

open source

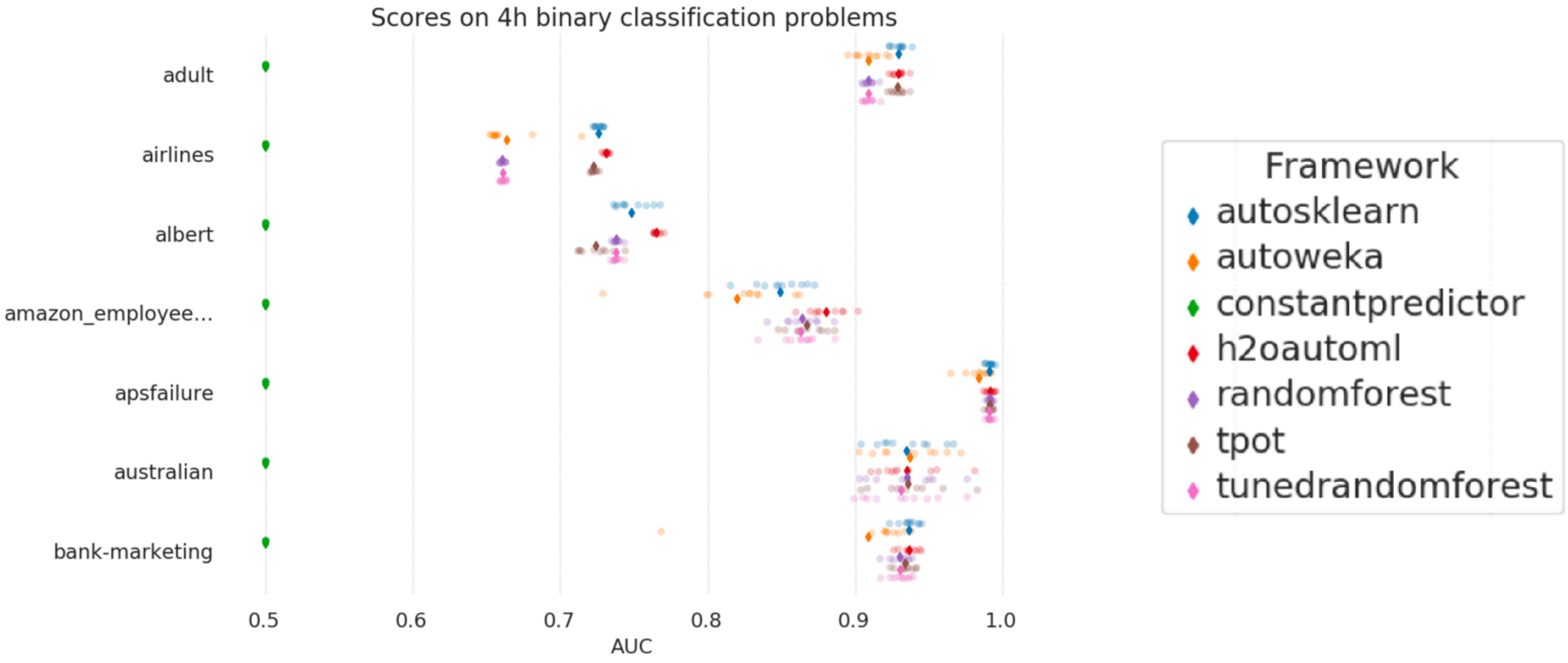
- AutoWEKA
- auto-sklearn
- TPOT
- H2O AutoML
- Auto-Keras
- Hyperopt-sklearn



Tool	Back-end	Optimization	Meta-learning	Post-processing
Auto-WEKA	WEKA	Bayesian	-	-
auto-sklearn	scikit-learn	Bayesian	warm-start	ensemble selection
TPOT	scikit-learn	Genetic Programming	-	-
H2O AutoML	H2O	Random Search	-	stacked ensembles

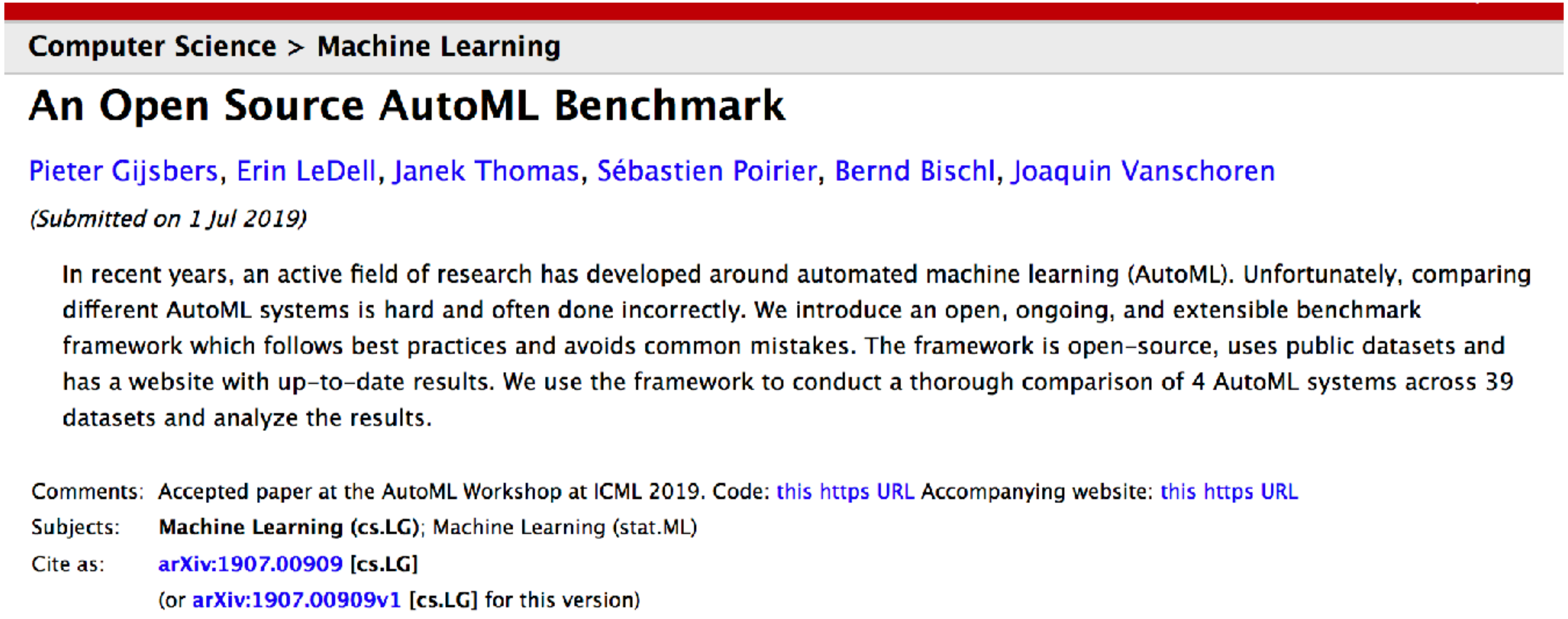
Table 1: Simplified comparison of a selection of AutoML tools.

AutoML Benchmark Results



https://openml.github.io/automlbenchmark/results.html

AutoML Benchmarks



Computer Science > Machine Learning

An Open Source AutoML Benchmark

Pieter Gijsbers, Erin LeDell, Janek Thomas, Sébastien Poirier, Bernd Bischl, Joaquin Vanschoren

(Submitted on 1 Jul 2019)

In recent years, an active field of research has developed around automated machine learning (AutoML). Unfortunately, comparing different AutoML systems is hard and often done incorrectly. We introduce an open, ongoing, and extensible benchmark framework which follows best practices and avoids common mistakes. The framework is open-source, uses public datasets and has a website with up-to-date results. We use the framework to conduct a thorough comparison of 4 AutoML systems across 39 datasets and analyze the results.

Comments: Accepted paper at the AutoML Workshop at ICML 2019. Code: this https URL Accompanying website: this https URL

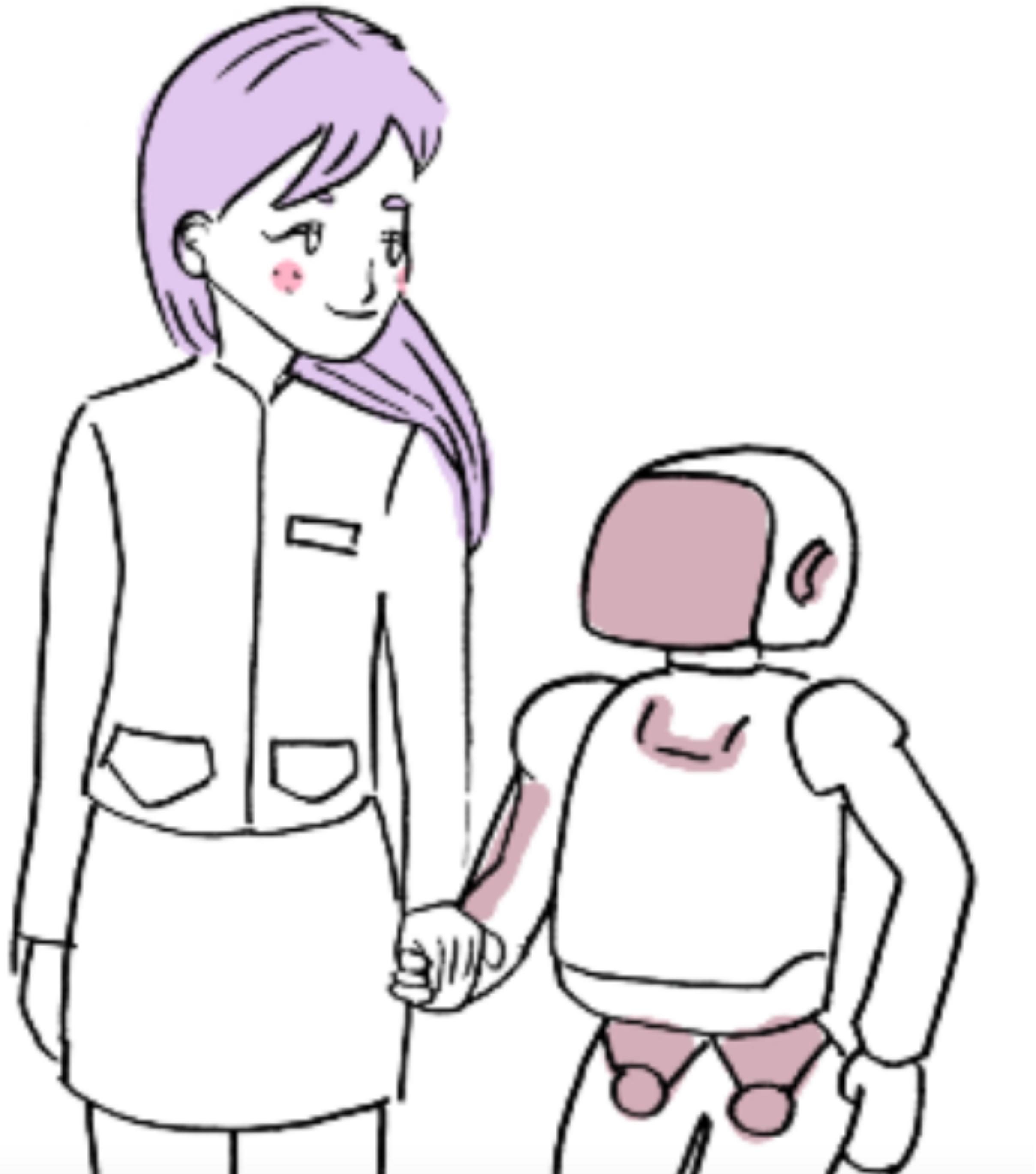
Subjects: **Machine Learning (cs.LG)**; Machine Learning (stat.ML)

Cite as: **arXiv:1907.00909 [cs.LG]**
(or **arXiv:1907.00909v1 [cs.LG]** for this version)

arXiv paper ⬇️

https://tinyurl.com/automlbenchmark

Thank you!



@ledell on Github, Twitter

erin@h2o.ai