A short story on AutoML

Katharina Eggensperger
katharina.eggensperger@uni-tuebingen.de

AutoML for Science
automl4science.de

September 2024

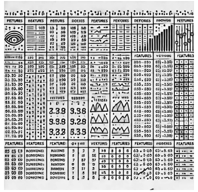
why & what

It’s awesome.
It’s complex.

Real-world workflows are challenging

Standard Workflow

Data

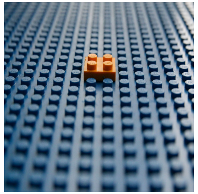


simplistic, unimodal and homogeneous

Approach

reasonable amounts of labelled data

Evaluation



validation on well-defined benchmarks

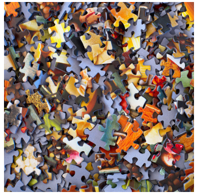
Setup

AI vs human comparison

New Challenges

noisy, multimodal, and heterogeneous

large unlabelled or small data situations

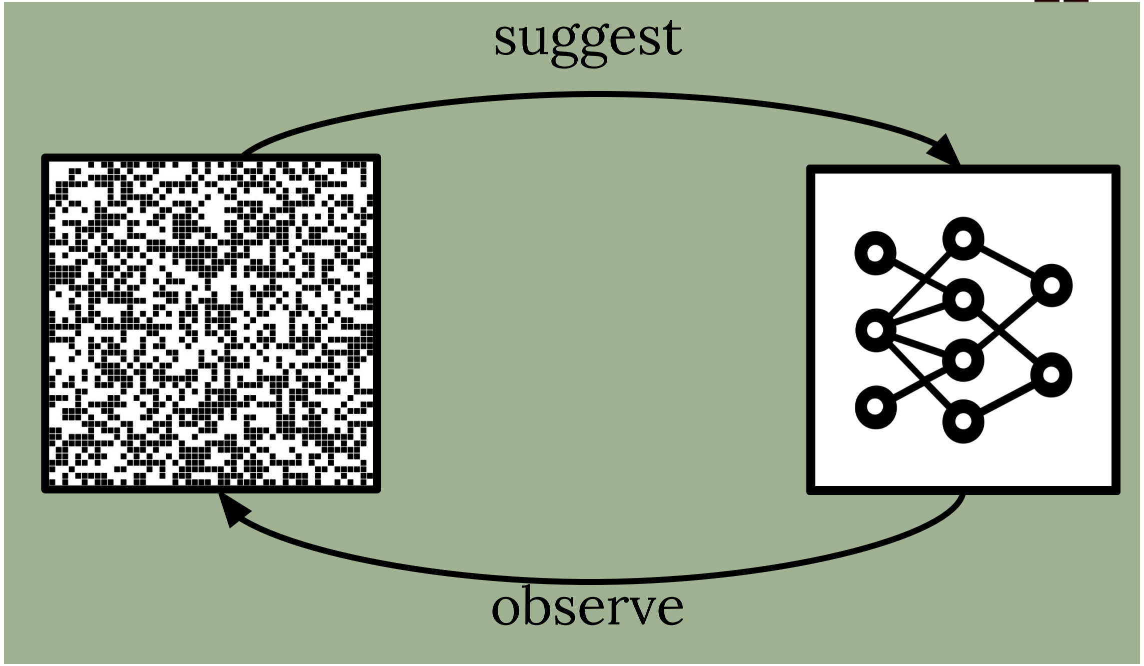


new settings

AI-human collaboration

Adapted from: P. Rajpurkar, E. Chen, O. Banerjee, E. Topol. *AI in health and medicine*. In: Nature Medicine 28 (2022).

Sequential Optimization



AutoML-assisted workflows

Goals

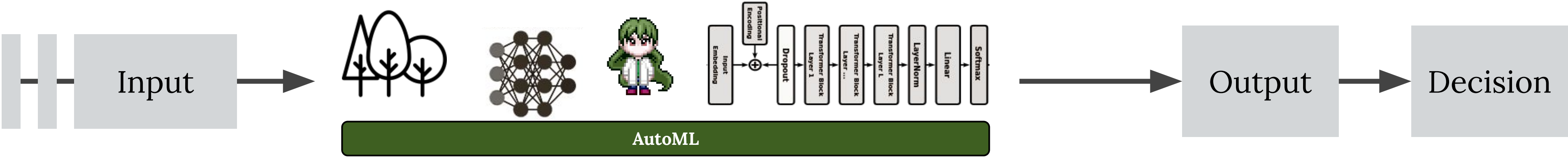
🦾⏱️ **high-performance** solutions

🔋🌳 **efficient** resources usage

Why?

🍏🍎 **fair** comparisons

🪄❓ **robust** development



A really short story on AutoML

$$\lambda^* \in \arg\min_{\lambda \in \Lambda} c(A_\lambda)$$

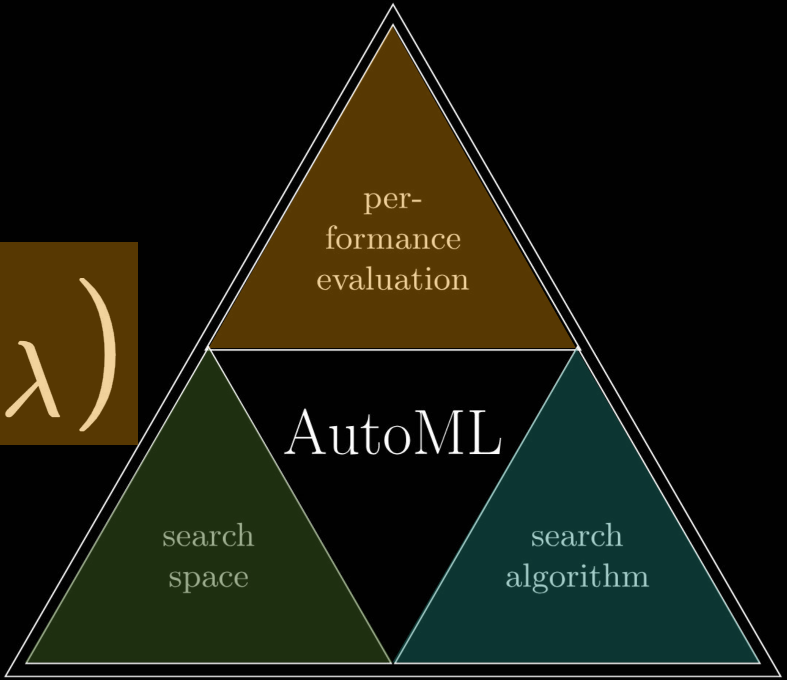


Image source: M. Baratchi, C. Wang, S. Limmer, J. van Rijn, H. Hoos, T. Bäck, M. Olhofer: *Automated machine learning: past, present and future*. *Artif Intell Rev* 57, 122 (2024).

Examples of AutoML Methods

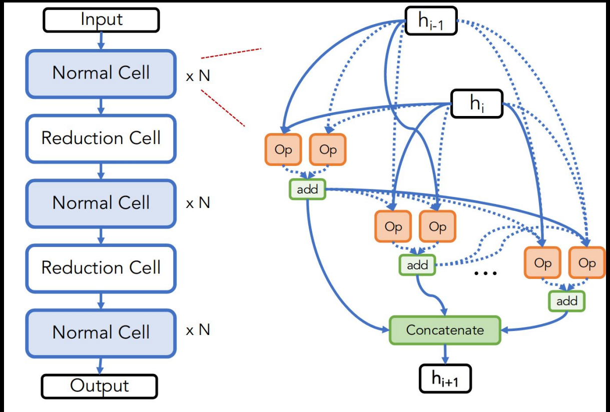


Image source: C. White, M. Safari, R. Sukthanker, B. Ru, T. Elsken, A. Zela, D. Dey, F. Hutter: *Neural Architecture Search: Insights from 1000 Papers* arXiv (2023).

Neural Architecture Search
What is the best DL architecture for my task (and hardware)?

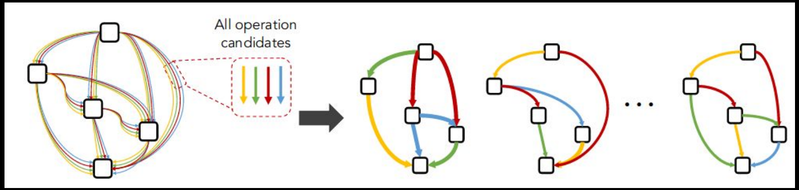
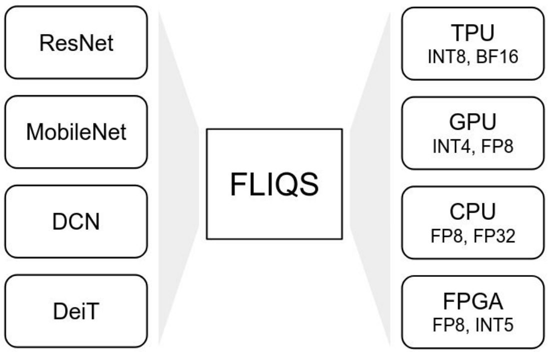


Image source: C. White, M. Safari, R. Sukthanker, B. Ru, T. Elsken, A. Zela, D. Dey, F. Hutter: *Neural Architecture Search: Insights from 1000 Papers* arXiv (2023).



← AutoML'24

J. Dotzel, G. Wu, A. Li, M. Umar, Y. Ni, M. Abdelfattah, Z. Zhang, L. Cheng, M. Dixon, N. Jouppi, Q. Le, S. Li: *FLIQS: One-Shot Mixed-Precision Floating-Point and Integer Quantization Search* In: AutoML Conf (2024)

Examples of AutoML Methods

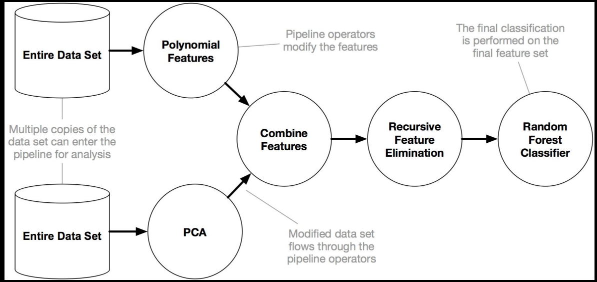


Image source: R. Olson. N. Bartley, R. Urbanowicz and J. Moore: *Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science.* In: ACM (2016)

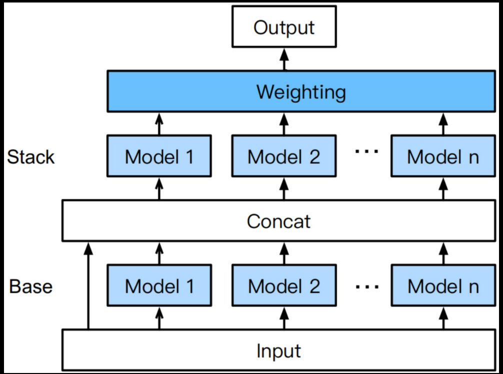


Image source: N. Erickson, J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, A. Smola: *AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data.* In: AutoML@ICML (2020)

AutoML Systems

What is the best ML pipeline for my task (and resource constraints)?

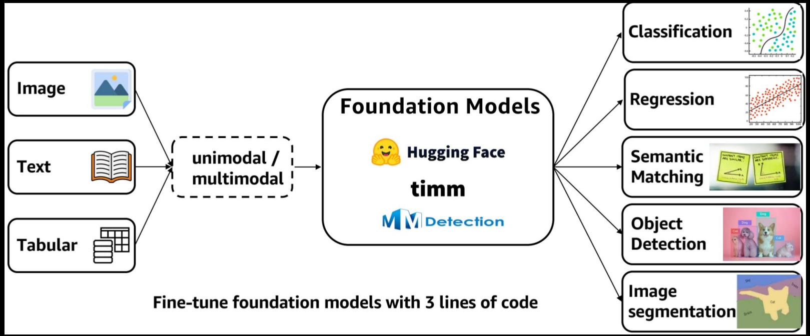


Image source: Zhiqiang Tang, Haoyang Fang, Su Zhou, Taojiannan Yang, Zihan Zhong, Cuixiong Hu, Katrin Kirchhoff, George Karypis: *AutoGluon-Multimodal (AutoMM): Supercharging Multimodal AutoML with Foundation Models.* In AutoML'24

← AutoML'24

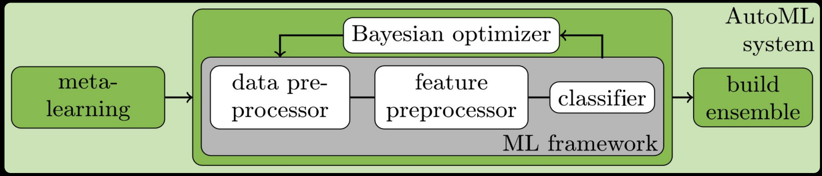
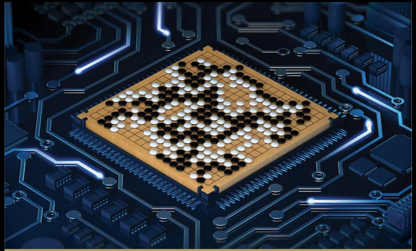


Image source: M. Feurer, A. Klein, **K. Eggensperger,** T. Springenberg, M. Blum, F. Hutter: *Efficient and Robust Automated Machine Learning.* In: NeurIPS 2015

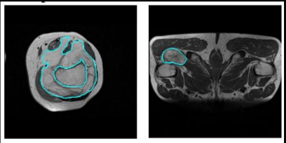
AutoML Applications

Game Playing



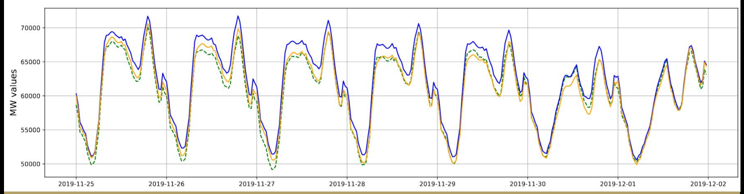
Y. Chen, A. Huang, Z. Wang, I. Antonoglou, J. Schrittwieser, D. Silver, N. de Freitas: *Bayesian Optimization in AlphaGo.* In: arXiv (2018)

Medical Research



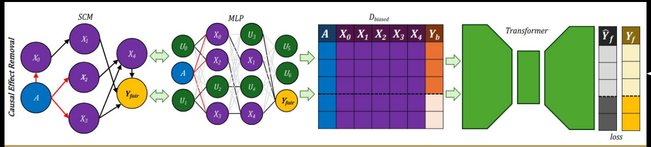
Martijn P.A. Starmansa et al. *Reproducible radiomics through automated machine learning validated on twelve clinical applications.* In: arXiv (2021)

Load Forecasting



J. Keisler, S. Claudel, G. Cabriel, M. Brégère: *Automated Deep Learning for load forecasting.* In: AutoML'24

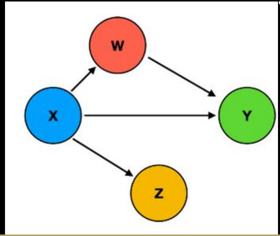
Algorithmic Fairness



J. Robertson, N. Hollmann, N. Awar, F. Hutter: *FairPFN: Transformers Can Do Counterfactual Fairness.* In: AutoML'24

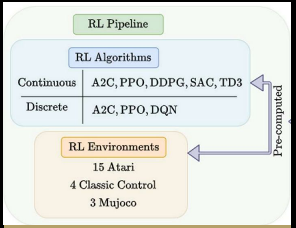
AutoML'24

Causal Discovery



G. Chan, T. Claassen, H. Hoos, T. Heskes, M. Baratchi: *AutoCD: Automated Machine Learning for Causal Discovery Algorithms.* In: AutoML (2024)

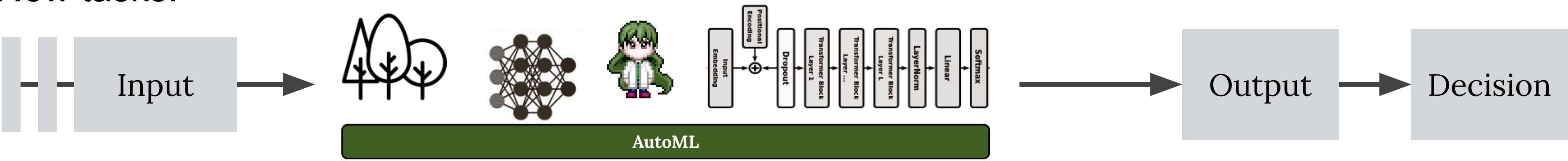
Reinforcement Learning



G. Shala, S. Arango, A. Biedenkapp, F. Hutter, J. Grabocka: *HPO-RL-Bench: A Zero-Cost Benchmark for HPO in Reinforcement Learning.* In: AutoML'24

(Auto-)ML assisted workflows

New tasks:



Standard tasks:

One use case: tabular data

- Simple and easy-accessible
- Available in many domains
- Challenging for ML

→ **High Demand for AutoML**

- HPO methods
- AutoML Systems for non-ML-experts

Culmen Length	Culmen Depth	Flipper Length	Weight	Sex	Species
39.1	18.7	181	3750	♂	Adelie
39.5	17.4	186	3800	♀	Adelie
40.3	18.0	195	3250	♀	Adelie
35.3	18.9	187	3800	♀	Adelie
40.6	18.6	183	3550	♂	Adelie
40.5	17.9	187	3200	♀	Adelie
42.3	21.2	191	4150	♂	Adelie
45.2	17.8	198	3950	♀	Chinstrap
46.1	18.2	178	3250	♀	Chinstrap
49.8	15.9	229	5950	♂	Gentoo
43.5	15.2	213	4650	♀	Gentoo
51.5	16.3	230	5500	♂	Gentoo
46.2	14.1	217	4375	♀	Gentoo
55.1	16.0	230	5850	♂	Gentoo

Features
(=what we observe)

Targets
(=what the model predicts)

train →

test ←

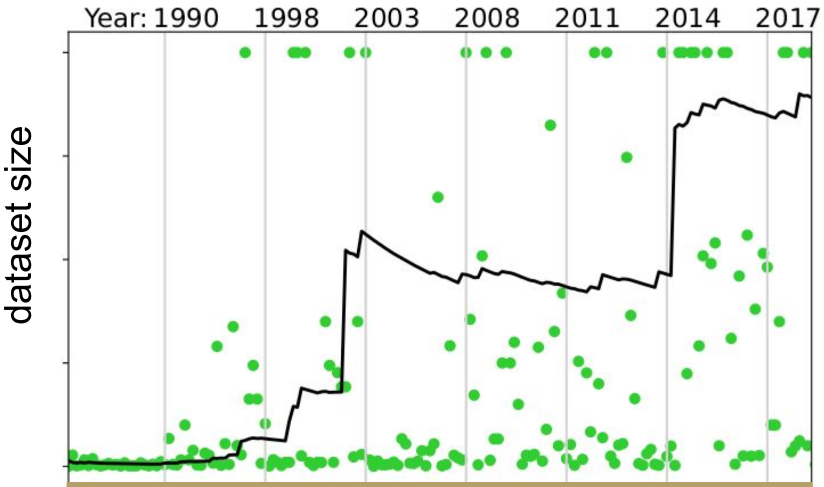
Culmen Length	Culmen Depth	Flipper Length	Weight	Sex
36.7	19.3	193	3450	♀
39.3	20.6	190	3650	♂
38.9	17.8	181	3625	♀
51.4	19	201	3950	♂
45.7	17.3	193	3600	♀
50.7	19.7	203	4050	♂
42.5	17.3	187	3350	♀
48.1	15.1	209	5500	♂
50.5	15.2	216	5000	♀

One use case: tabular data

- Simple and easy-accessible
- Available in many domains
- Challenging for ML

→ **High Demand for AutoML**

- HPO methods
- AutoML Systems for non-ML-experts



R. Kohli, M. Feurer, **K. Eggensperger**, B. Bischl, F. Hutter *Towards Quantifying the Effect of Datasets for Benchmarking: A Look at Tabular Machine Learning* In: DMLR@ICLR'24

Let’s build an AutoML system

Should work on **any** supervised tabular classification task.

Standardize in- and output

Data, Metric, Budget → ? → Predictions, Model

Should work on **any** supervised tabular classification task.

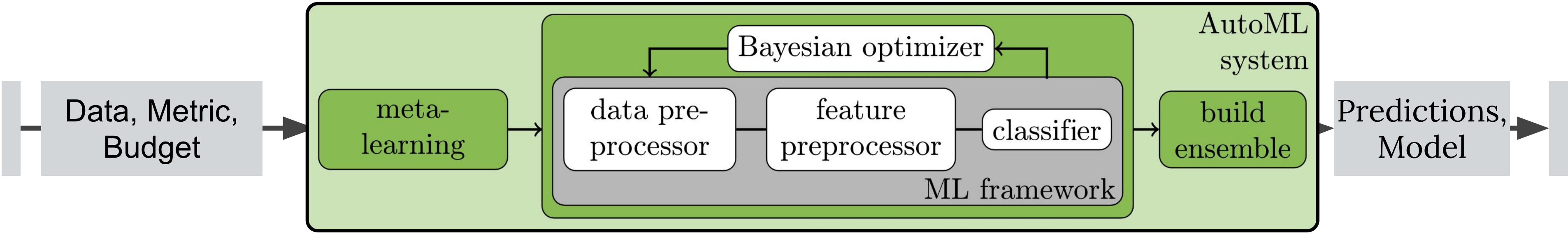
Needed: ML in four lines of code

Data, Metric, Budget

```
import automl
>>> cls = automl.Classifier()
>>> cls.fit(X_train, y_train)
>>> predictions = cls.predict(X_test)
```

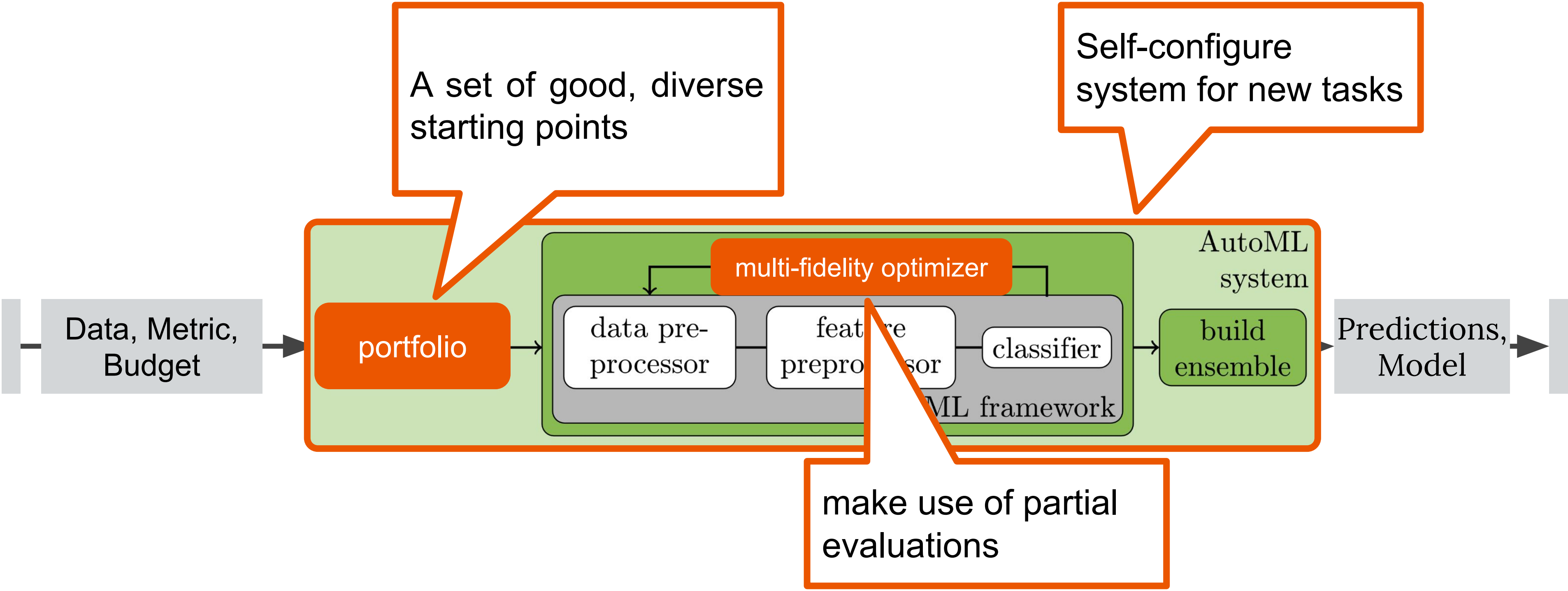
Predictions, Model

One solution: Auto-sklearn 1.0



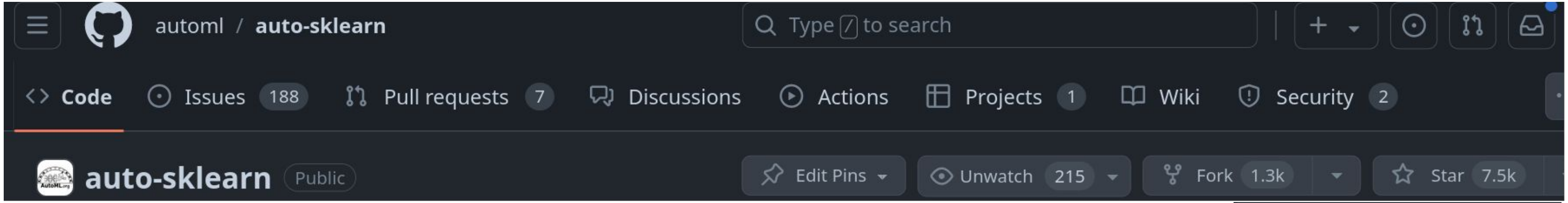
M. Feurer, A. Klein, **K. Eggensperger,** T. Springenberg, M. Blum, F. Hutter: Efficient and Robust Automated Machine Learning. In: NeurIPS 2015

Auto-sklearn 1.0 to Auto-sklearn 2.0

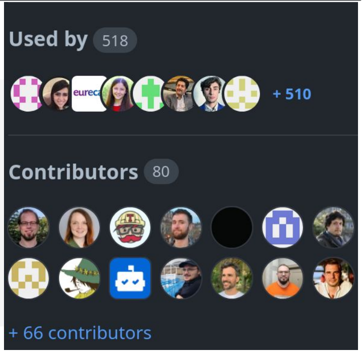


M. Feurer, A. Klein, **K. Eggensperger,** T. Springenberg, M. Blum, F. Hutter: Efficient and Robust Automated Machine Learning. In: NeurIPS 2015
M. Feurer, **K. Eggensperger**, S. Falkner, M. Lindauer, F. Hutter: "Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning. In: JMLR 2022

Open-source Auto-sklearn




```
import autosklearn.classification
>>> cls = autosklearn.classification.AutoSklearnClassifier()
>>> cls.fit(X_train, y_train)
>>> predictions = cls.predict(X_test)
```



- used for many **applications** (>**2.5K** citations **and >15K** downloads/month)
- **won** 2 AutoML **competitions**

/automl/auto-sklearn
/automl/auto-sklearn-talks

M. Feurer, A. Klein, **K. Eggensperger**, T. Springenberg, M. Blum, F. Hutter: Efficient and Robust Automated Machine Learning. In: NeurIPS 2015
M. Feurer, **K. Eggensperger**, S. Falkner, M. Lindauer, F. Hutter: “Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning. In: JMLR 2022

An (incomplete) timeline of AutoML systems for tabular data

2004 — Ensemble selection from libraries of models [Caruana et al.]

2005 — Gene Expression Model Selection [Statnikov et al.]

NeurIPs Agnostic Learning vs Prior Learning 2007

2010 — Ensemble particle swarm model selection [Escalante et al]

2012 — GA-PSO-FMS [Sun et al]; Auto-WEKA [Thornton et al.]

2014 — Hyperopt-sklearn [Komer et al]

2015 — Auto-sklearn [Feurer et al]

AutoML I 2015/16

2016 — TPOT [Olson et al]; H20 [LeDell et al]

AutoML II 2017/18

AutoML III 2018

2019 — Auto-Keras [Jin et al]

2020 — Auto-Gluon [Erickson et al]

2021 — FLAML [Wang et al]; LightAutoML [Vakhrushev et al]

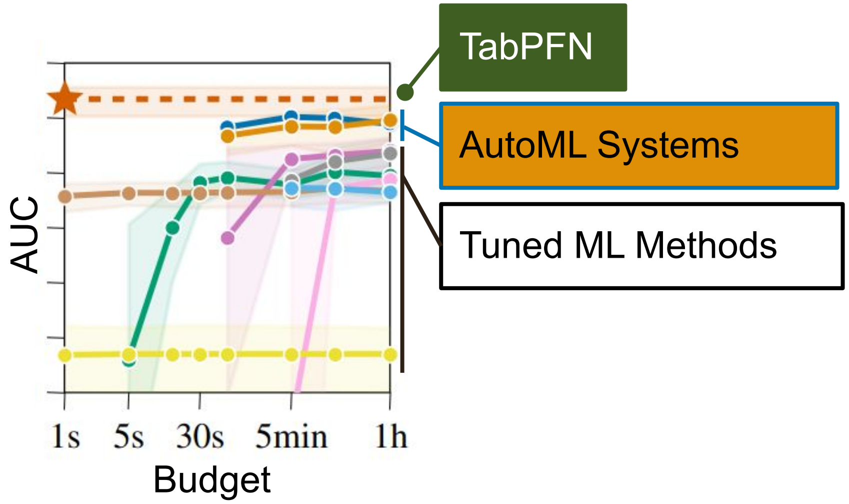
2022 — Auto-sklearn 2.0 [Feurer et al]

AutoML Benchmak 2022

AutoML GrandPrix 2024

Tabular (Auto-)ML now. ICL?

Data, Metric, Budget → **TabPFN (in-context learning)** → Predictions / Model



TL;DR TabPFN, a transformer pre-trained on synthetic data, that instantly yields predictions for tabular datasets.

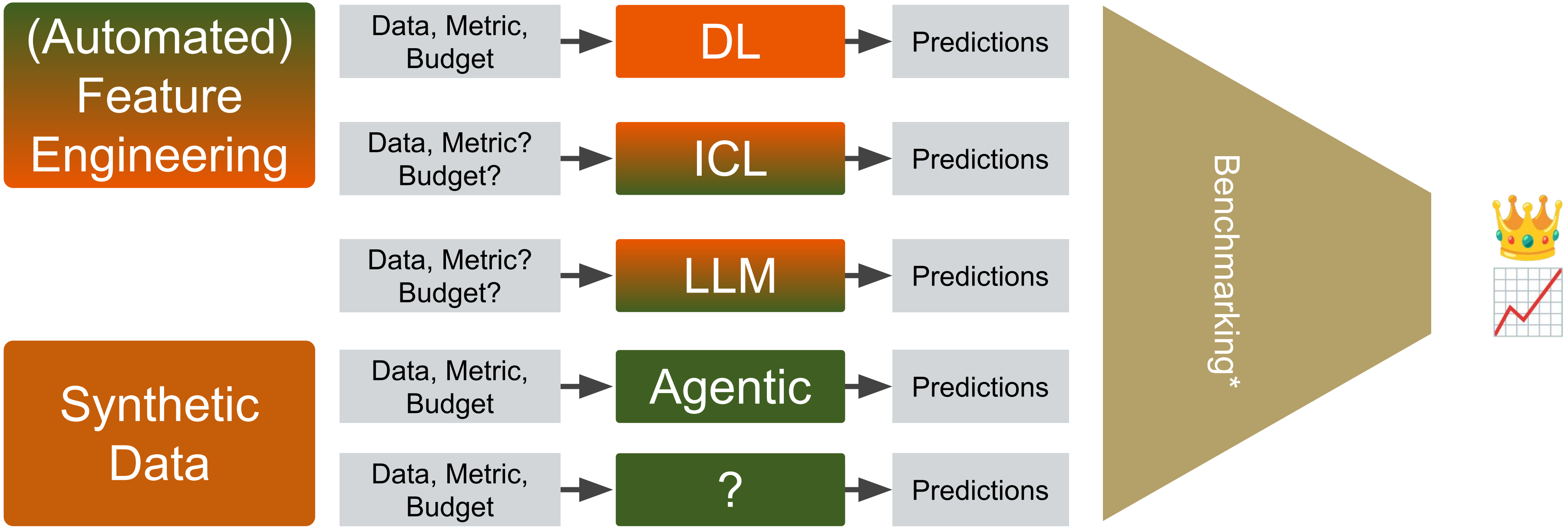
Remarks:

- Up to 1000 samples, 100 features, 10 classes
- works best on **continuous** datasets **w/o missing** values

/automl/TabPFN

N. Hollmann, S. Müller, **K. Eggensperger**, F. Hutter: “TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second”. In: ICLR'23

Tabular (Auto-)ML now. LLMs? Agentic Data Science? Architectures? Synthetic Data Generation?



*See also M. Hardt: *The emerging science of benchmarks.* Invited Talk at ICLR’24

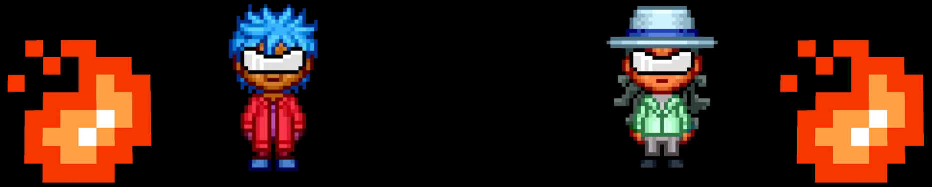
...the end? 👑📈

Or not?

$c(A)$

IT'S DANGEROUS TO GO ALONE!

TAKE THIS. AND THIS.



Λ

$\arg\min$

$c(A)$

When applying **(Auto-)ML**

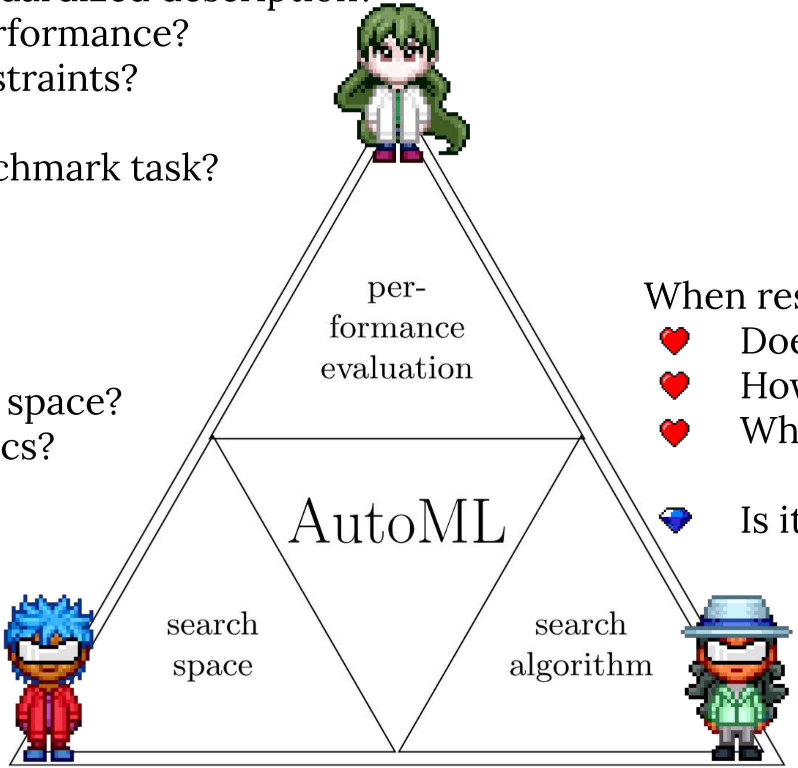
- ❤ Can we frame it in a standardized description?
- ❤ How can we measure performance?
- ❤ Are there additional constraints?

- 💎 Could this be an ML benchmark task?

When researching **ML methods**

- ❤ Is the method flexible?
- ❤ What's an effective search space?
- ❤ Are there priors / heuristics?

- 💎 Could this be an AutoML benchmark task?



When researching **AutoML methods**

- ❤ Does it scale to real-world ML?
- ❤ How does it perform across tasks?
- ❤ Where does it not work?

- 💎 Is it easy-to-use?

Image source: M. Baratchi, C. Wang, S. Limmer, J. van Rijn, H. Hoos, T. Bäck, M. Olhofer: *Automated machine learning: past, present and future. Artif Intell Rev* 57, 122 (2024).

ML needs AutoML[1]
AutoML needs ML[2]

[1] AutoML makes progress in ML accessible.
[2] ML is the foundation of AutoML.

Thanks!

Katharina Eggensperger
katharina.eggensperger@uni-tuebingen.de
automl4science.de