

# Benchmarki

Struktura artykułu naukowego

# Benchmarki



# Nie istnieje jeden ustalony system benchmarków

- algorytmy ML
- taski (dla danych tabelarycznych zbiory danych)
- metryki

Ćwiczenie 1: <https://arxiv.org/pdf/1907.00909.pdf>

Ćwiczenie 2: <https://arxiv.org/pdf/1802.09596.pdf>



# Don't trust any ML benchmark that:

1. Does not split training and test data
2. Does not use the identical train/test split(s) across each algorithm being tested
3. Does not do multiple train/test splits
4. Uses less than five different datasets (or appropriately qualifies its conclusions)
5. Uses less than three well-established metrics to evaluate the results (or appropriately qualifies its conclusions)
6. Relies on one of the services / software packages being tested to compute quality metrics
7. Does not use the same software to compute all quality metrics for all algorithms being tested

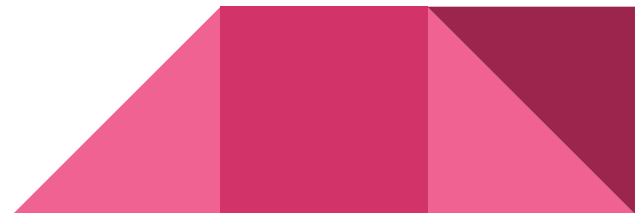


# AutoML challenges

## Analysis of the AutoML Challenge Series 2015–2018

- **time limit: 20 min**
- **Codalab**
- **Different data distributions:** the intrinsic/geometrical complexity of the dataset.
- **Different tasks:** regression, binary classification, multi-class classification, multi-label classification.
- **Different scoring metrics:** AUC, BAC, MSE,  $F_1$ , etc. (see Sect. 10.4.2).
- **Class balance:** Balanced or unbalanced class proportions.
- **Sparsity:** Full matrices or sparse matrices.
- **Missing values:** Presence or absence of missing values.
- **Categorical variables:** Presence or absence of categorical variables.
- **Irrelevant variables:** Presence or absence of additional irrelevant variables (distractors).
- **Number  $P_{tr}$  of training examples:** Small or large number of training examples.
- **Number  $N$  of variables/features:** Small or large number of variables.
- **Ratio  $P_{tr}/N$  of the training data matrix:**  $P_{tr} \gg N$ ,  $P_{tr} = N$  or  $P_{tr} \ll N$ .

# AutoML challenges - phases

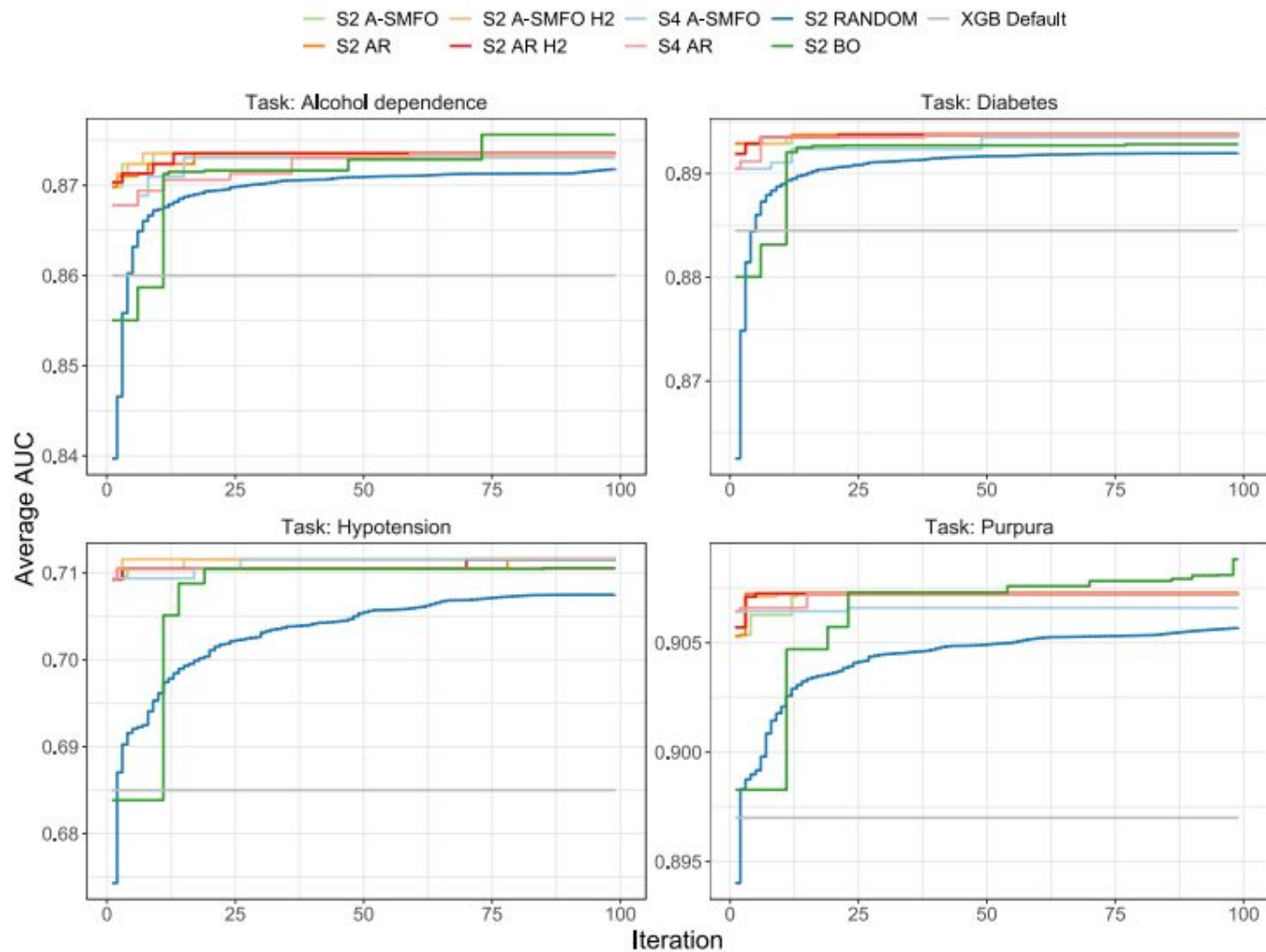


# AutoML challenges - results

2. AutoML phase				1. Feedback phase			
Ended	Winners	< $R$ >	< $S$ >	Ended	Performance	< $R$ >	< $S$ >
03/31/18	1. aad_freiburg	2.80	0.4341	03/12/18	aad_freiburg	9.0	0.7422
	2. narnars0	3.80	0.4180		narnars0	4.40	0.7324
	3. wlWangl	5.40	0.3857		wlWangl	4.40	0.8029
	3. thanhng	5.40	0.3874		thanhng	14.0	0.6845
	3. Malik	5.40	0.3863		Malik	13.8	0.7116

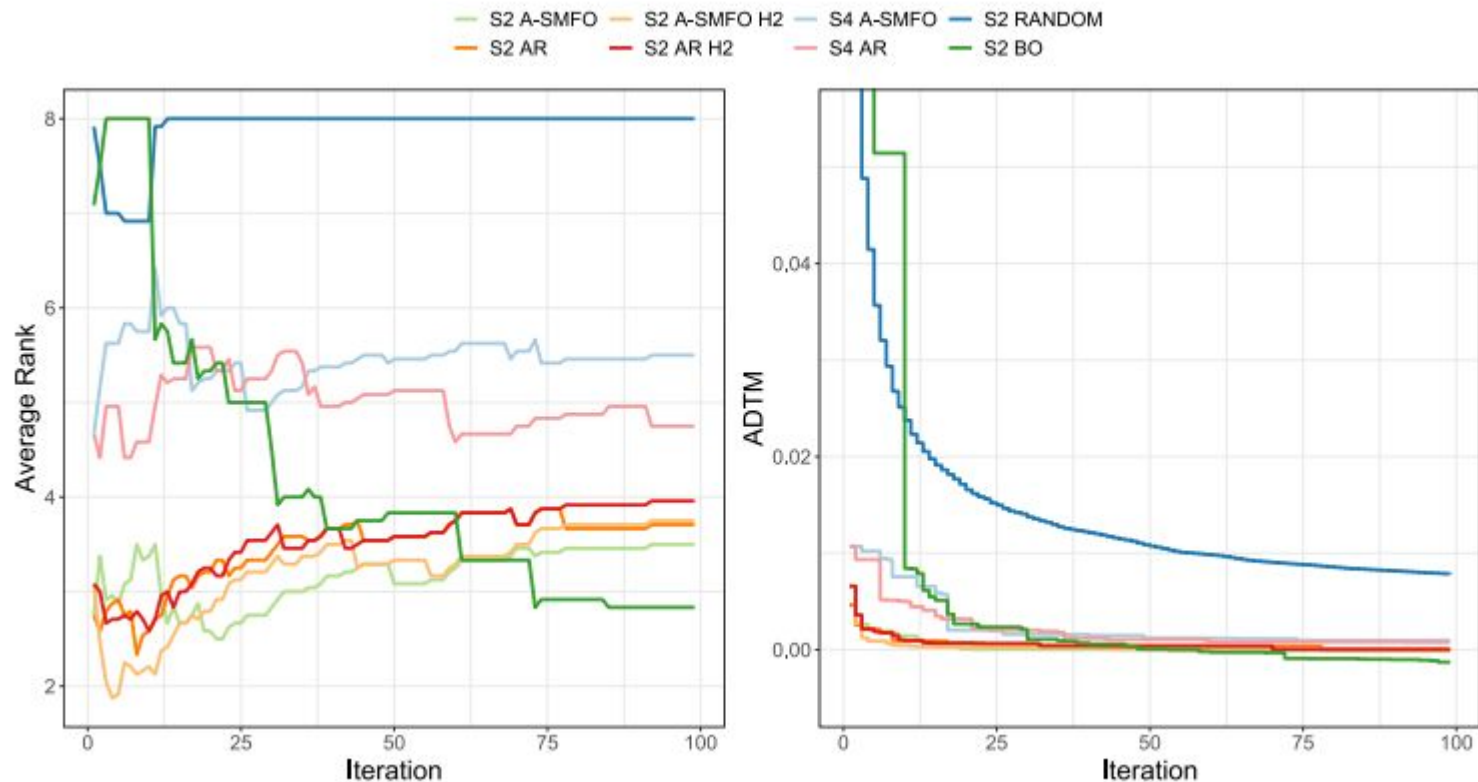
Table 10.6 Results of the 2018 challenge winners. Each phase was run on five different datasets. We show the winners of the AutoML (blind test) phase and for comparison their performances in the Feedback phase.

# Anytime performance





# Agregacja wyników



## Average distance to minimum/maximum

$$\text{ADTM}(\Lambda_T, \mathcal{D}) := \frac{1}{|\mathcal{D}|} \sum_{D \in \mathcal{D}} \min_{\lambda \in \Lambda_T} \frac{f_D(\lambda) - f_D^{\min}}{f_D^{\max} - f_D^{\min}}$$



# Struktura artykułu

Title, Author,  
Abstract,  
Keywords

- Descriptive information that lets readers search for an article.

Introduction

- What is the context for this project?
- How does it fit in with other research on the topic?
- *What is the research question?*

**WHY?**

Methods

- What did the author(s) do to answer the research question?

**HOW?**

Results

- What was the answer to the question?
- This is often shown in tables and figures.

**WHAT?**

Discussion/  
Conclusion

- What is the significance of this project?
- How does it fit in with what else is known about the topic?

**SO WHAT?**

References

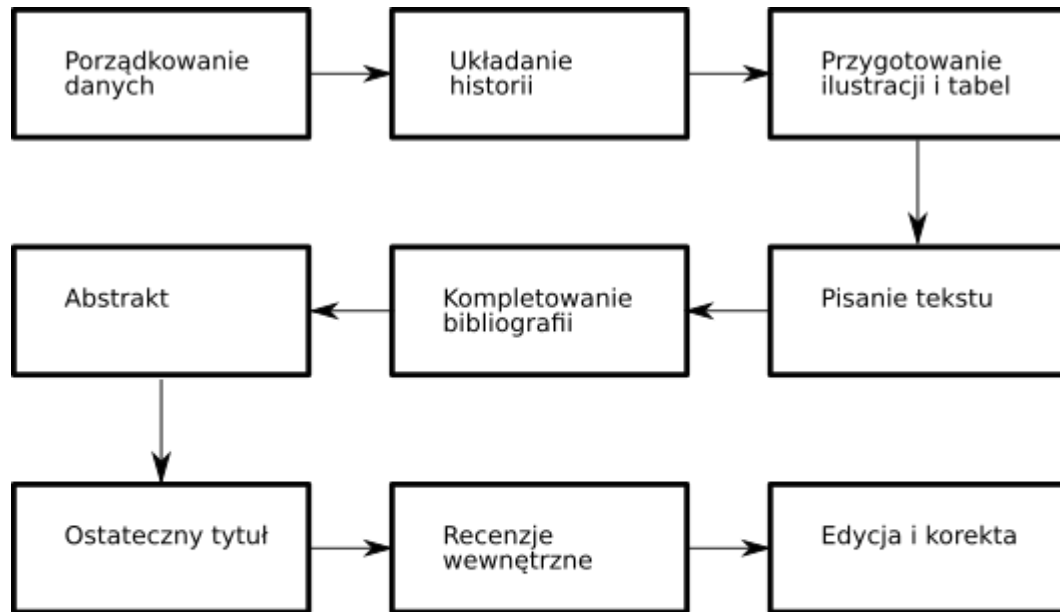
- Materials the author(s) cited when writing this paper.

<https://library.sumdu.edu.ua/en/for-researcher/academic-writing/writing-scientific-text/the-structure-of-scientific-texts/823-the-structure-of-scientific-texts.html>

# Kolejność pisania sekcji artykułu



# Kolejność pisania sekcji artykułu



# Kolejność pisania sekcji artykułu

1. Przygotowanie do publikacji - generowanie wyników + przygotowanie tła naukowego np. zapisywanie referencji
  - zapisywanie wszystkich ustawień aparatury
2. Rezultaty (można przygotować sobie pierwszą wersję abstraktu)
3. Metody
4. Dyskusja
5. Wprowadzenie
6. Abstrakt (ostateczny)
7. Tytuł (ostateczny)



# Methods - cele sekcji

- dokładnie opisujemy jak przeprowadzaliśmy nasze badania, z jakich metod korzystaliśmy
- pokazanie poprawności i wiarygodności naszych badań
- reprodukowalność

Warto podawać tabele z różnymi ustawieniami, linki do githuba. Często bardzo dokładne opisy są przenoszone do appendix-u.





# Rezultaty

- **prezentujemy główne** wyniki badań
- kluczowa jest selekcja najważniejszych informacji dotyczących się pytania badawczego
- oprócz *suchych* wyników trzeba zadbać o zrozumiały opis
- sekcja powinna być zbudowana jako historia przeprowadzająca czytelnika od najważniejszych rezultatów do mniej istotnych



# Rezultaty - prezentacja wyników

- narracja powinna być zbudowana wokół tabel i wykresów
- wszystkie zaprezentowane wyniki powinny być pokazane w jakimś celu, należy się do nich odwołać w tekście (nie musimy do każdej zaprezentowanej wartości)
- w tekście nie powinniśmy się skupiać na wartościach liczbowych ale na ich omówieniu

czy wartość  $u = 2.04$  to jest dużo czy mało? czy potwierdza jakieś przypuszczenia?



# Wykresy i tabele

- **czy lepiej przedstawić dane jako tabela czy jako wykres?**
- podpisy pod tabelami/wykresami: co jest tam przedstawione , jak czytać dany wykres
- warto zadbać o jednolity styl
- odnoszenie się w tekście:

Odpowiedzi kobiet były zbliżone do odpowiedzi mężczyzn (wykres 2).

Wykres 2 pokazuje, że odpowiedzi kobiet i mężczyzn były podobne.



Ćwiczenie: <https://arxiv.org/pdf/1802.09596.pdf>

Figure 1 - jak jest opisane w tekście?

Table 3 + Figure 3 - jak się uzupełniają?

Figure 2 - caption

Ćwiczenie: <https://arxiv.org/pdf/1907.00909.pdf>

Table 2 czy Figure 1 - co jest lepsze?



Title, Author,  
Abstract,  
Keywords

- Descriptive information that lets readers search for an article.

Introduction

- What is the context for this project?
- How does it fit in with other research on the topic?
- *What is the research question?*

WHY?

Methods

- What did the author(s) do to answer the research question?

HOW?

Results

- What was the answer to the question?
- This is often shown in tables and figures.

WHAT?

Discussion/  
Conclusion

- What is the significance of this project?
- How does it fit in with what else is known about the topic?

SO WHAT?

References

- Materials the author(s) cited when writing this paper.

# Budowanie historii

# Wprowadzenie - cele

1. pokazanie dlaczego ten artykuł powstał? (purpose statement)  
dlaczego ten artykuł chce wypełnić lukę w badaniach?  
dlaczego ta luka wymaga wypełnienia?
2. Osadzenie artykułu w szerszym kontekście (background and context)  
zapewnienie informacji wejściowych wystarczających aby czytelnik mógł  
ocenić istotność przedstawionych badań



# Wprowadzenie - struktura

1. Pole badań - opis powinien uświadamiać dlaczego to problem jest ważny
2. Aspekty poruszanego problemu
3. Jaka jest nisza badawcza w poruszonym problemie?
4. Cele badania - jak zostanie wypełniona ta nisza?

Wprowadzenie nie musi być długie ani składać się z jednej sekcji ale musi dawać dobre podsumowanie kontekstu prac.

Artykuł z czasopisma: [Tunability](#)

Artykuł z konferencji: <https://arxiv.org/pdf/1905.12982.pdf>

<https://ml.informatik.uni-freiburg.de/papers/15-AAAI-MI-SMBO.pdf>



# Referencje

- pokazują co zostało zrobione wcześniej, czytelnik może odnieść się do opisanych wcześniej zagadnień
- pokazujemy, że nie tylko my się tym problemem zajmowaliśmy, podkreślamy jego wagę
- zwiększamy wiarygodność naszych badań - wcześniej zrobiliśmy przegląd literatury

Powinniśmy odnosić się głównie do research articles





# Cytowania

- skupione na informacjach (information prominent focus)

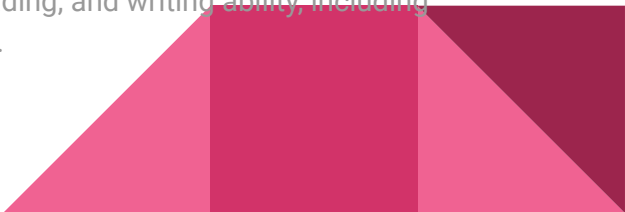
Plagiarism is a mounting concern in higher education and professional circles (Bertram Gallant, 2011; Dee & Jacob, 2010; McCabe, 2005).

- skupione na autorze (author prominent focus)

Bergstra and Bengio (2012) compute the relevance of the hyperparameters of neural networks and conclude that some are important on all datasets, while others are only important on some datasets.

- mieszane (weak author prominent focus)

Several researchers have examined the relationship between plagiarism, academic reading, and writing ability, including summarizing. . . (Howard, 2008; Howard, Serviss, & Rodrigue, 2010; Pennington, 2010).



# Dyskusja - cele

- interpretacja rezultatów w świetle dotychczas znanych wyników (nawiązanie do wprowadzenia)
- rozbudowana odpowiedź na pytanie badawcze: jak zmienił się stan wiedzy względem początku artykułu
- podkreślenie głównych wniosków z pracy: take home message



# Elementy kluczowe dla widoczności

# Rola tytułu, abstraktu, słów kluczowych

1. w wyszukiwarkach artykuły są pozycjonowane na podstawie tych elementów
2. widoczne dla wszystkich, nawet w bazach z ograniczonym dostępem
3. są pierwszymi elementami z którym zapoznają się recenzenci i czytelnicy  
powinny więc zachęcać do przeczytania

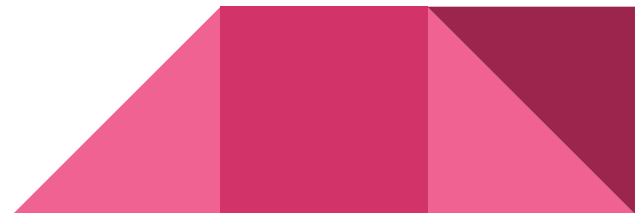


# Tytuł- strategie

1. Zwięzły tytuł informacyjny zawierający najważniejsze odkrycie lub pole badań
2. Użycie jak największej ilości słów kluczowych - ważna jest lokalizacja słów w tytule
3. Tytuł w formie zdania oznajmującego, zbudowane wokół głównego rzeczownika (przykład *Light-Driven Soft Robot Mimics Caterpillar Locomotion in Natural Scale*)
4. Tytuł dwuczłonowy (przykład *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*)



1. *Tunability: Importance of Hyperparameters of Machine Learning Algorithms*
2. *Using Meta-Learning to Initialize Bayesian Optimization of Hyperparameters*
3. *Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning*
4. *Attention Is All You Need*
5. *All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously*
6. *From Fourier to Koopman: Spectral Methods for Long-term Time Series Prediction*



# Struktura abstraktu

- |      |  |   |            |
|------|--|---|------------|
| I.   | Część kontekstowa - dziedzina badań i jej znaczenie                        | } | 1-3 zdania |
| II.  | Główne pytanie badawcze - jaką lukę w dotychczasowej wiedzy chcę wypełnić? |   |            |
| III. | Mój oryginalny wkład - co nowego wniosły moje badania?                     | } | 3-5 zdań   |
| IV.  | Część metodologiczna - co i jak badałem/badałam?                           |   |            |
| V.   | Rezultaty - jakie są główne wyniki?  |   | 3-5 zdań   |
| VI.  | Dyskusja - co z tego wynika, take home message                             |   | 1-3 zdania |



# Efficiently Sampling Functions from Gaussian Process Posteriors,

by James T. Wilson, Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, Marc Peter Deisenroth

Gaussian processes are the gold standard for many real-world modeling problems, especially in cases where a model's success hinges upon its ability to faithfully represent predictive uncertainty. These problems typically exist as parts of larger frameworks, wherein quantities of interest are ultimately defined by integrating over posterior distributions. These quantities are frequently intractable, motivating the use of Monte Carlo methods. Despite substantial progress in scaling up Gaussian processes to large training sets, methods for accurately generating draws from their posterior distributions still scale cubically in the number of test locations. We identify a decomposition of Gaussian processes that naturally lends itself to scalable sampling by separating out the prior from the data. Building off of this factorization, we propose an easy-to-use and general-purpose approach for fast posterior sampling, which seamlessly pairs with sparse approximations to afford scalability both during training and at test time. In a series of experiments designed to test competing sampling schemes' statistical properties and practical ramifications, we demonstrate how decoupled sample paths accurately represent Gaussian process posteriors at a fraction of the usual cost.





# Tunability: Importance of Hyperparameters of Machine Learning Algorithms

by Philipp Probst, Anne-Laure Boulesteix, Bernd Bischl

Modern supervised machine learning algorithms involve hyperparameters that have to be set before running them. Options for setting hyperparameters are default values from the software package, manual configuration by the user or configuring them for optimal predictive performance by a tuning procedure. The goal of this paper is two-fold. Firstly, we formalize the problem of tuning from a statistical point of view, define data-based defaults and suggest general measures quantifying the tunability of hyperparameters of algorithms. Secondly, we conduct a large-scale benchmarking study based on 38 datasets from the OpenML platform and six common machine learning algorithms. We apply our measures to assess the tunability of their parameters. Our results yield default values for hyperparameters and enable users to decide whether it is worth conducting a possibly time consuming tuning strategy, to focus on the most important hyperparameters and to choose adequate hyperparameter spaces for tuning.

