

# Wprowadzenie do komunikacji naukowej

Struktura artykułów cz. 2

# Podstawowe zasady pisania artykułów naukowych

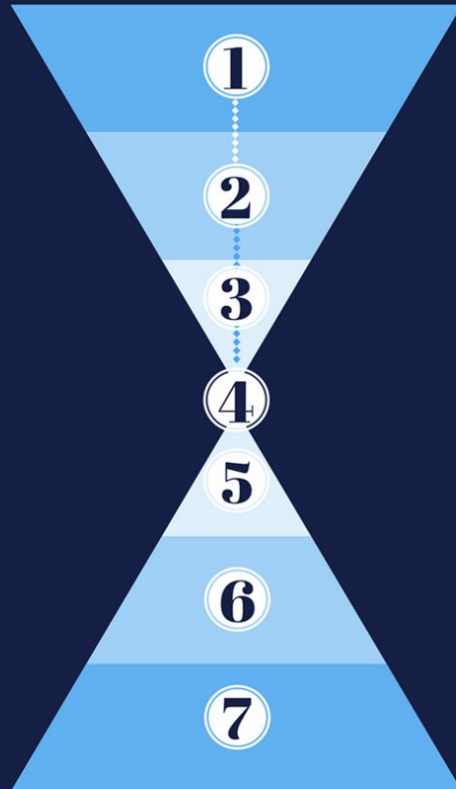
- Opis badań nie jest historią, którą chcesz opowiedzieć
- Wyniki Twoich badań nie są historią, którą chcesz opowiedzieć

Wszystkie wyniki, które pokażemy w artykule powinny mieć na celu odpowiedź na pytanie: Co wynika z badań i jakie mają znaczenie?

PYTANIE BADAWCZE



# THE ANATOMY OF A SCIENTIFIC PAPER



**1** INTRODUCTION:  
What is known?  
(Our understanding of the world)

**2** What is unknown?  
(What's the gap we want to fill?)

**3** How and why  
should we fill the gap?  
(Your rationale and purpose/hypothesis)

**4** METHODS:  
What did you do?

**5** RESULTS:  
What results did you get?

**6** DISCUSSION:  
How do the results fill the gap?

**7** CONCLUSION:  
What does this mean for us going forward?

Learn more at [wordvice.com](https://wordvice.com)

<https://wordvice.com/seminar-how-to-write-an-effective-research-paper/>

# 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

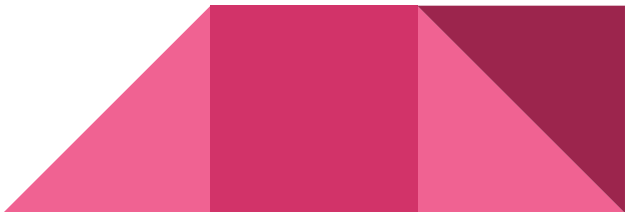
1. Pole badań - opis powinien uświadamiać dlaczego to problem jest ważny
2. Aspekty poruszanego problemu
3. Jaka jest nisza badawcza w poruszonym problemie?
4. Cele badania - jak zostanie wypełniona ta nisza?

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)

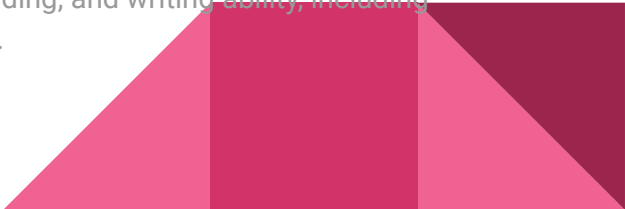
Plagiarism is a mounting concern in higher education and professional circles (Bertram Gallant, 2011; Dee & Jacob, 2010; McCabe, 2005).

- 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).



# Cytowania

- autor-data

(Eiben and Smit, 2011)

(Guyon et al., 2010)

- numerowane

Automated Machine Learning (AutoML) [20] is an emerging field...

(...) the review by Feurer and Hutter [11].





# Metody - cele

- 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.



# Metody

- opis eksperymentu, który został wykonany np. opis aparatury lub ustawień oprogramowania [tu](#)
- opis jak zbierane były dane
- opis teoretyczny np. definicje, dotychczas udowodnione twierdzenia [tu](#) [tu](#)
- opis metod, które zostały zaimplementowane w oprogramowaniu [tu](#)



# Rezultaty

- prezentujemy główne wyniki badań
- kluczowa jest selekcja najważniejszych informacji dotyczących 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?

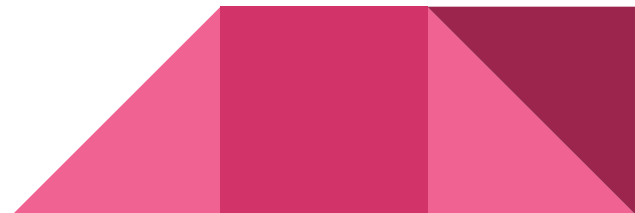


# Wykresy i tabele

- 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.



Parameter	Def.P	Def.O	Tun.P	Tun.O	$q_{0.05}$	$q_{0.95}$
glmnet			0.042	0.019		
alpha	1	0.252	0.022	0.010	0.015	0.979
lambda	0	0.005	0.029	0.017	0.001	0.223
rpart			0.020	0.012		
cp	0.01	0.002	0.013	0.008	0	0.528
maxdepth	30	19	0.004	0.004	10	28
minbucket	7	5	0.005	0.006	1.85	43.15
minsplit	20	13	0.002	0.003	6.7	47.6
kknn			0.021	0.008		
k	7	14	0.021	0.008	2	30
svm			0.041	0.030		
kernel	radial	radial	0.019	0.018		
cost	1	936.982	0.019	0.003	0.025	943.704
gamma	$1/p$	0.002	0.024	0.020	0.007	276.02
degree	3	3	0.005	0.014	2	4
ranger			0.016	0.007		
num.trees	500	162	0.001	0.001	203.5	1908.25
replace	TRUE	FALSE	0.004	0.001		
sample.fraction	1	0.76	0.003	0.003	0.257	0.971
mtry	$\sqrt{p}$	$p \cdot 0.432$	0.010	0.003	0.081	0.867
respect.unordered.factors	TRUE	TRUE	0.001	0.000		
min.node.size	1	1	0.001	0.002	0.009	0.453
xgboost			0.034	0.011		
nrounds	500	3342	0.004	0.002	1360	4847.15
eta	0.3	0.031	0.005	0.005	0.002	0.445
subsample	1	0.89	0.003	0.002	0.555	0.964
booster	gbtree	gbtree	0.008	0.005		
max_depth	6	14	0.001	0.001	3	13
min_child_weight	1	1.264	0.009	0.002	1.061	7.502
colsample_bytree	1	0.712	0.005	0.001	0.334	0.887
colsample_bylevel	1	0.827	0.006	0.001	0.348	0.857
lambda	1	2.224	0.002	0.002	0.004	5.837
alpha	1	0.021	0.003	0.002	0.003	2.904

Table 7: Tunability measures for single hyperparameters and tuning spaces as in Table 3, but calculated for the accuracy. Defaults (package defaults (Def.P) and own optimal defaults (Def.O)), tunability of the hyperparameters with the package defaults (Tun.P) and our optimal defaults (Tun.O) as reference and tuning space quantiles ( $q_{0.05}$  and  $q_{0.95}$ ) for different parameters of the algorithms.

**Table 1.** Accuracies and ranks (in parentheses) of the algorithms 1-nearest neighbor (IB1), C4.5 (J48), RIPPER (JRip), LogisticDiscriminant (LogD), MultiLayerPerceptron (MLP) and naive Bayes (NB) on different datasets and their mean rank.

Datasets	IB1	J48	JRip	LogD	MLP	NB
abalone	.197 (5)	.218 (4)	.185 (6)	.259 (2)	.266 (1)	.237 (3)
acetylation	.844 (1)	.831 (2)	.829 (3)	.745 (5)	.609 (6)	.822 (4)
adult	.794 (6)	.861 (1)	.843 (3)	.850 (2)	.830 (5)	.834 (4)
...	...	...	...	...	...	...
Mean rank	4.05	2.73	3.17	3.74	2.54	4.78

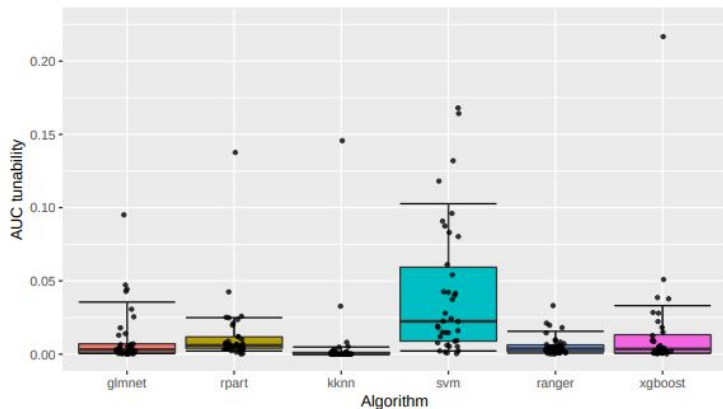
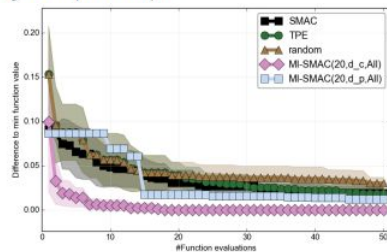
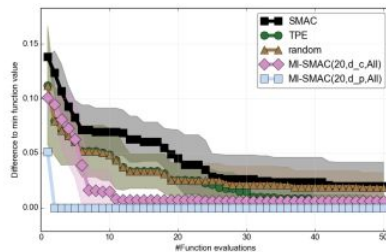


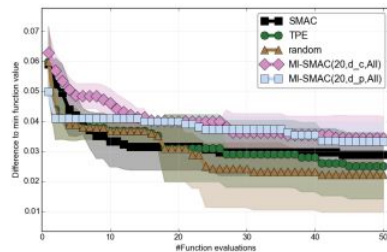
Figure 2: Boxplots of the tunabilities (AUC) of the different algorithms with respect to optimal defaults. The upper and lower whiskers (upper and lower line of the boxplot rectangle) are in our case defined as the 0.1 and 0.9 quantiles of the tunability scores. The 0.9 quantile indicates how much performance improvement can be expected on at least 10% of datasets. One outlier of `glmnet` (value 0.5) is not shown.



dataset: liver-disorders



dataset: heart-h



dataset: hepatitis

Figure 1: Difference in validation error between hyperparameters found by SMBO and the best value obtained via full grid search for three datasets with scikit-learn. (20, $d$ ,X) stands for MI-SMAC with an initial design of  $t = 20$  configurations suggested by meta-learning with distance measure  $d$  using metafeatures X.

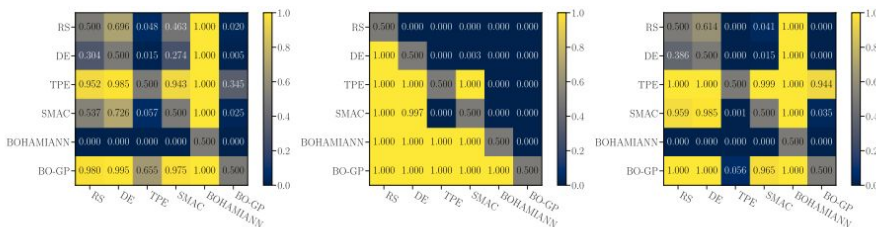


Figure 4: Heatmaps of the p-values of the pairwise Mann-Whitney U test on three scenarios. Small p-values should be interpreted as finding evidence that the method in the column outperforms the method in the row. Using tasks from our meta-model lead to results that are close to using the large set of original tasks. *Left*: results with 1000 real tasks. *Middle*: subset of only 9 real tasks. *Right*:

# Graficzny abstrakt

- jedna ilustracja podsumowująca rezultaty artykułu (graficzne take home message)
- np. wizualna reprezentacja metody lub schemat technologiczny
- miniaturka artykułu na stronie konferencji czy blogu

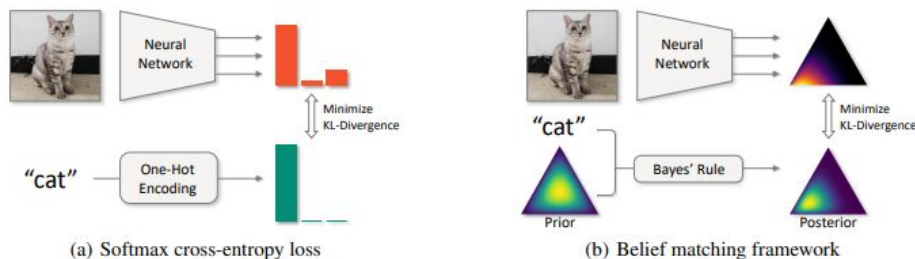


Figure 1. Illustration of the difference between softmax cross-entropy loss and belief matching framework when each image is unique in the training set. In softmax cross-entropy loss, the label “cat” is directly transformed into the target categorical distribution. In belief matching framework, the label “cat” is combined with the prior Dirichlet distribution over the categorical probability. Then, the Bayes’ rule updates the *belief* about categorical probability, which produces the target distribution.



# Dyskusja - struktura

- przegląd najważniejszych rezultatów: czy potwierdzają hipotezę badawczą, czy potwierdzają realizację założonych celów
- odniesienie się do innych prac: czy wnioski są podobne czy inne
- pokazanie ograniczeń przedstawionych badań
- implikacje wyników badań w szerszym kontekście
- rekomendacje dalszych prac

Warto zwrócić uwagę na siłę argumentacji zastosowaną w tej części



- 1) ... it appears that the existence of strong ties has the capacity to ...
- 2) These results indicate that the existence of strong ties will be determined by the cooperation ...
- 3) The presence of X suggests that the existence of strong ties may be influenced by the ...
- 4) Our results demonstrate that cooperation of members indicates the existence of strong ties...



# 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



# Praca domowa - struktura

- Zlokalizujcie w tekście sekcje wprowadzenia, metod, rezultatów i dyskusji. Czy są ułożone standardowo? Jeśli nie to zastanówcie się czemu miał służyć taki zabieg
- Czy sekcje są podzielone na mniejsze podsekcje? Jak konstruowane są podtytuły?
- Przy pomocy jakich środków zostały pokazane rezultaty? Jak opisane są wykresy i tabele?

Czy jest coś co Was zaskoczyło w układzie artykułu?



# Praca domowa - językowa

- zaznaczcie po dwa cytowania skupione na informacjach, autorze i mieszane
- jaki czas został zastosowany we wprowadzeniu Present Simple czy Present Perfect?
- W sekcji metody i rezultaty dominuje strona bierna czy czynna? Znajdźcie przykłady zastosowania strony biernej i czynnej

