ChaLearn AutoML challenges

Hugo Jair Escalante, Isabelle Guyon, The ChaLearn collaboration

The AAAI 2019 Workshop on Reproducible AI, Honolulú, HI, January, 27, 2019

Contents

- Challenges in ML
- ChaLearn AutoML challenges
- Discussion

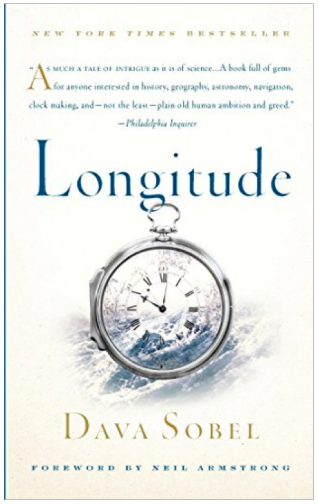
Challenges in ML

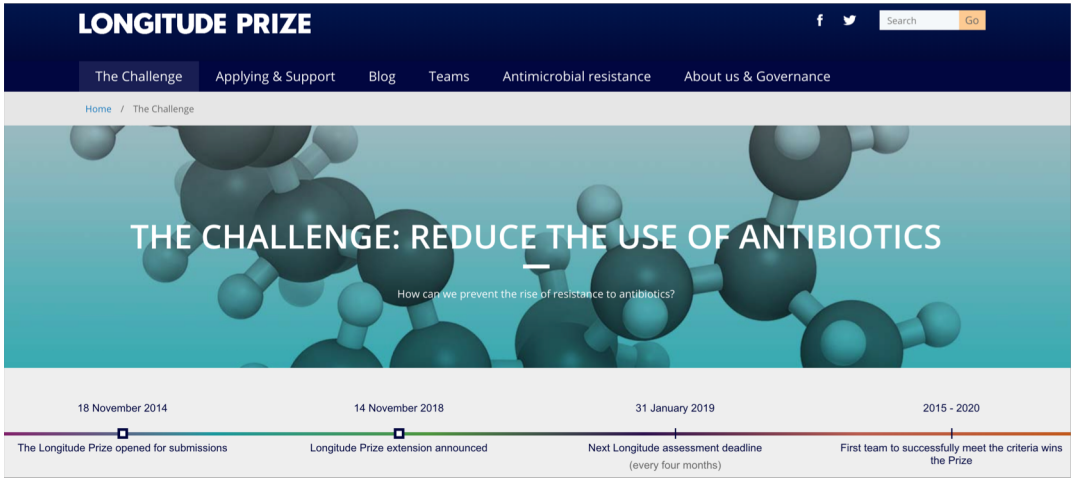
Challenges

- An effective mechanism to rapidly advance the state of the art and solve *practical* problems
  - Organized around a specific and well defined problem/task of scientific or practical relevance
  - Organizers provide problem/task, data, evaluation protocol, rules, prizes, platform and dissemination channels
  - Challenges can be industry or scientifically oriented

Challenges

- An effective mechanism to rapidly advance the state of the art and solve *practical* problems





https://longitudeprize.org/challenge

Challenges

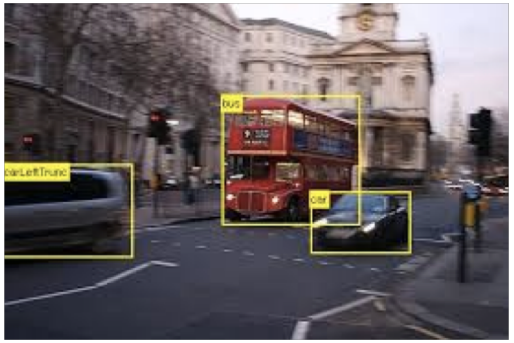
- An effective mechanism to rapidly advance the state of the art and solve *practical* problems

https://www.netflixprize.com/

Challenges

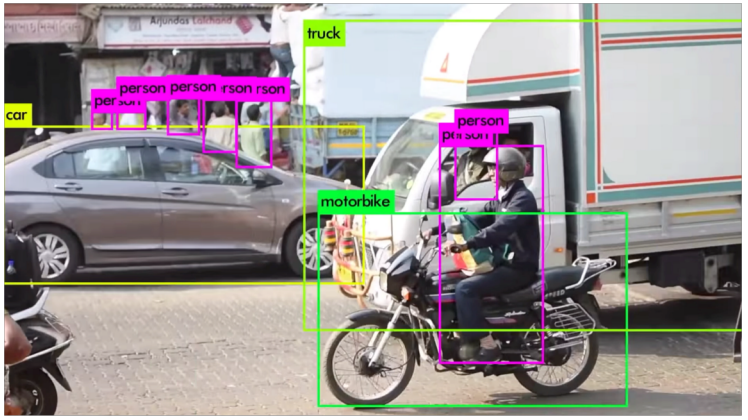
- An effective mechanism to rapidly advance the state of the art and solve *practical* problems

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning



http://host.robots.ox.ac.uk/pascal/VOC/

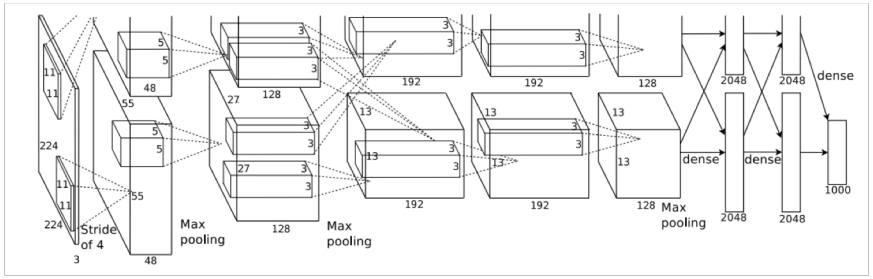
YOLO



https://pjreddie.com/darknet/yolo/

Challenges

- An effective mechanism to rapidly advance the state of the art and solve *practical* problems

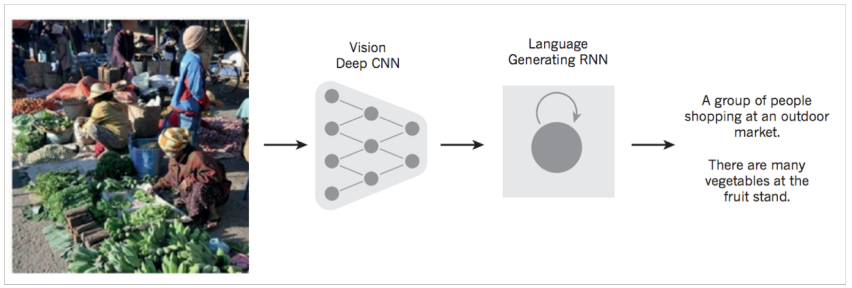


Russakovsky, O., Deng, J., Su, H. et al. **ImageNet Large Scale Visual Recognition Challenge.** Int J Comput Vis (2015) 115: 211.
A Krizhevsky, I Sutskever, GE Hinton . **Imagenet classification with deep convolutional neural networks.** NIPS 2012, 1097-1105

Challenges

- An effective mechanism to rapidly advance the state of the art and solve *practical* problems

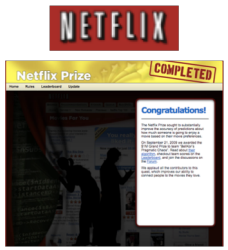
http://cocodataset.org/



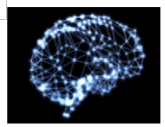
https://pdollar.wordpress.com/2015/01/21/image-captioning/

Challenges

- An effective mechanism to rapidly advance the state of the art and solve *practical* problems

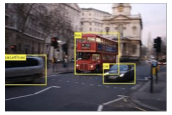


Higgs challenge



COCO Common Objects in Context

PASCAL2 Pattern Analysis, Statistical Modelling and Computational Learning



CLEF INITIATIVE The CLEF Initiative Conference and Labs of the Evaluation Forum

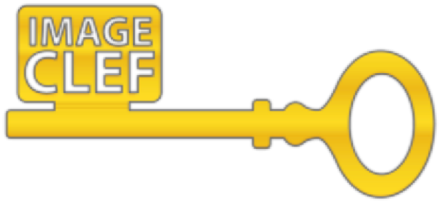
Challenges

- Pros
  - Problems of practical importance are solved (to some extend) rapidly (collaborative-competitive problem solving)
  - Contribute to establish benchmarks and the fair comparison among methodologies **(reproducibility)**
  - Participants are (self-) trained on new *practical* skills
  - Promotes team work

- Cons
  - Organization of a successful challenge is not an easy task (problem formulation, data gathering, evaluation protocols, funding, platform ….)
  - Challenge may degenerate into inflexible solutions / extremely complex

Challenge organization

- Platforms, organizations!



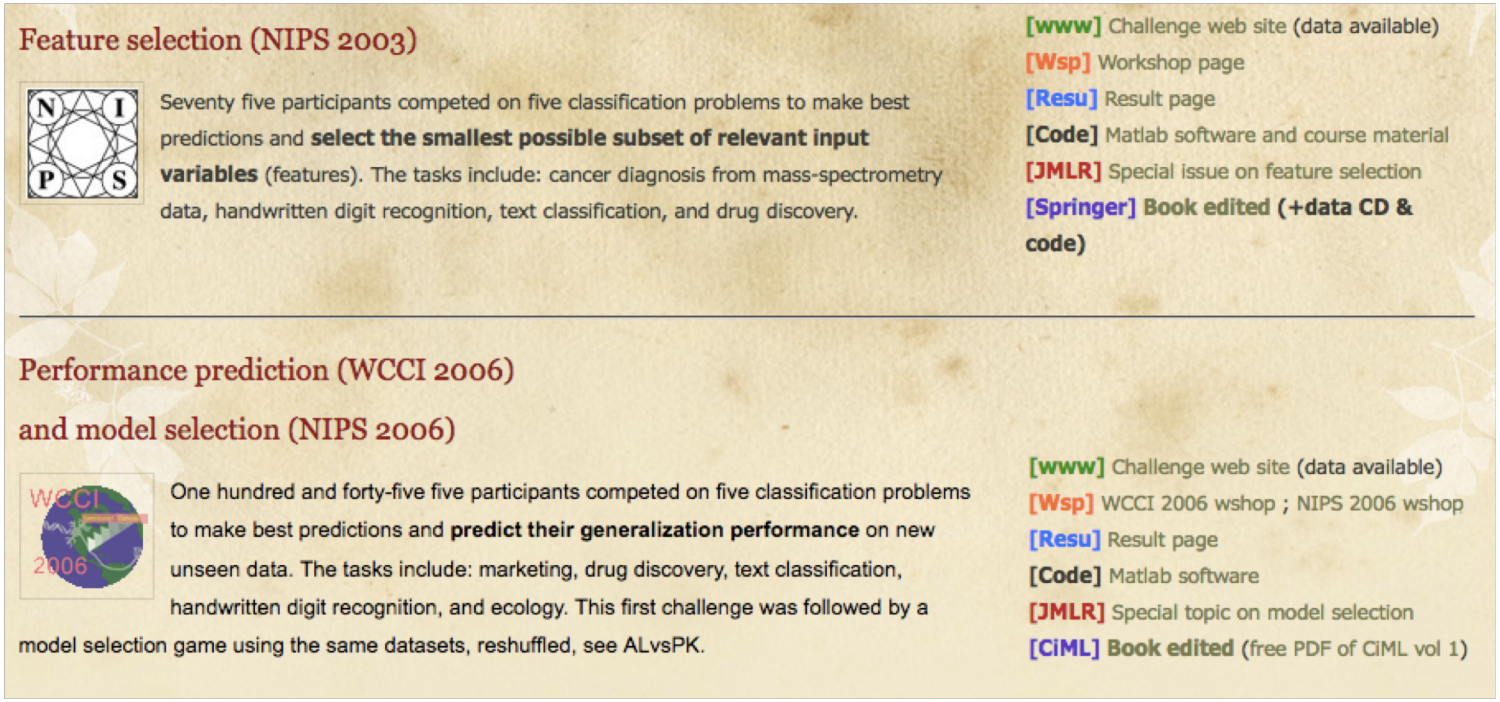
CodaLab

ChaLearn

- ChaLearn is a non-profit organization focusing on organization of challenges in Machine Learning (and related fields)

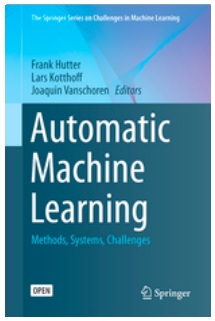
http://chalearn.org

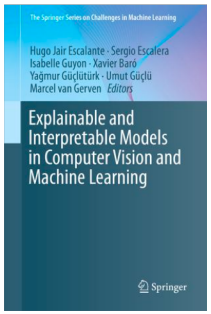
ChaLearn

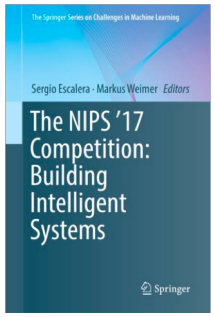


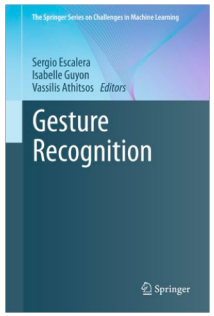
http://chalearn.org

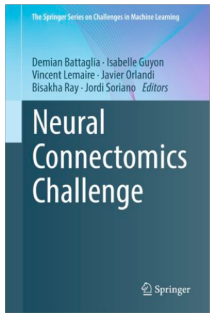
Springer Series on CiML



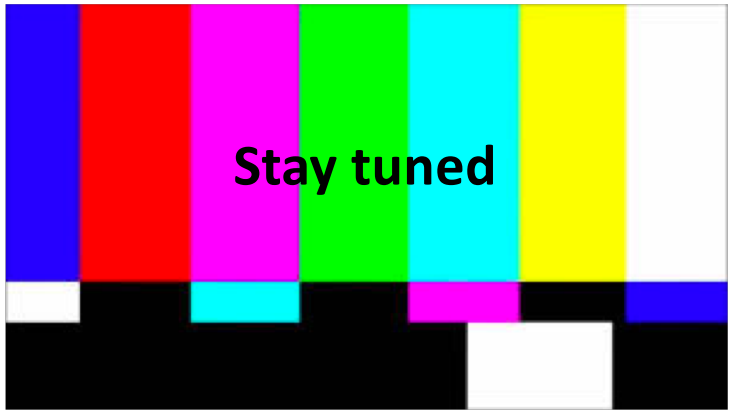








http://www.springer.com/series/15602



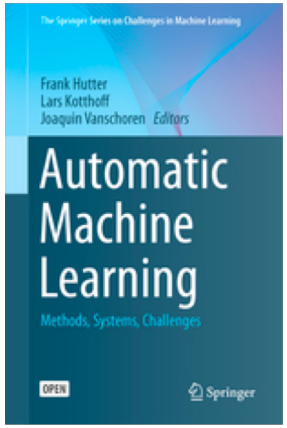
http://chalearn.org/

ChaLearn AutoML Challenges

AutoML

- Automatic Machine Learning*
  - Research area that targets progressive automation of machine learning
  - Field of research focusing on the development of autonomous methods for solving a variety of machine learning problems

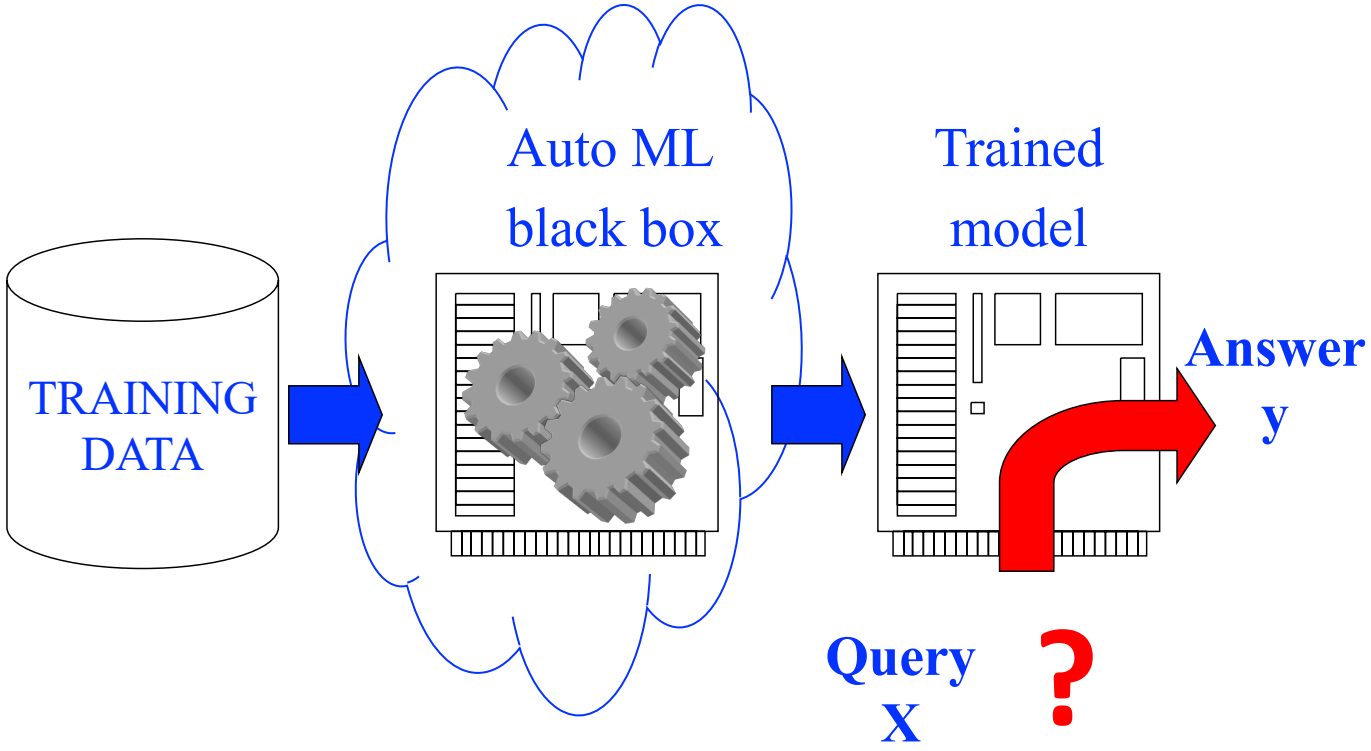
- Motivation.
  - Large amounts of data readily available everywhere
  - Lack of domain and/or ML experts who can advise/supervise the development of ML-based systems



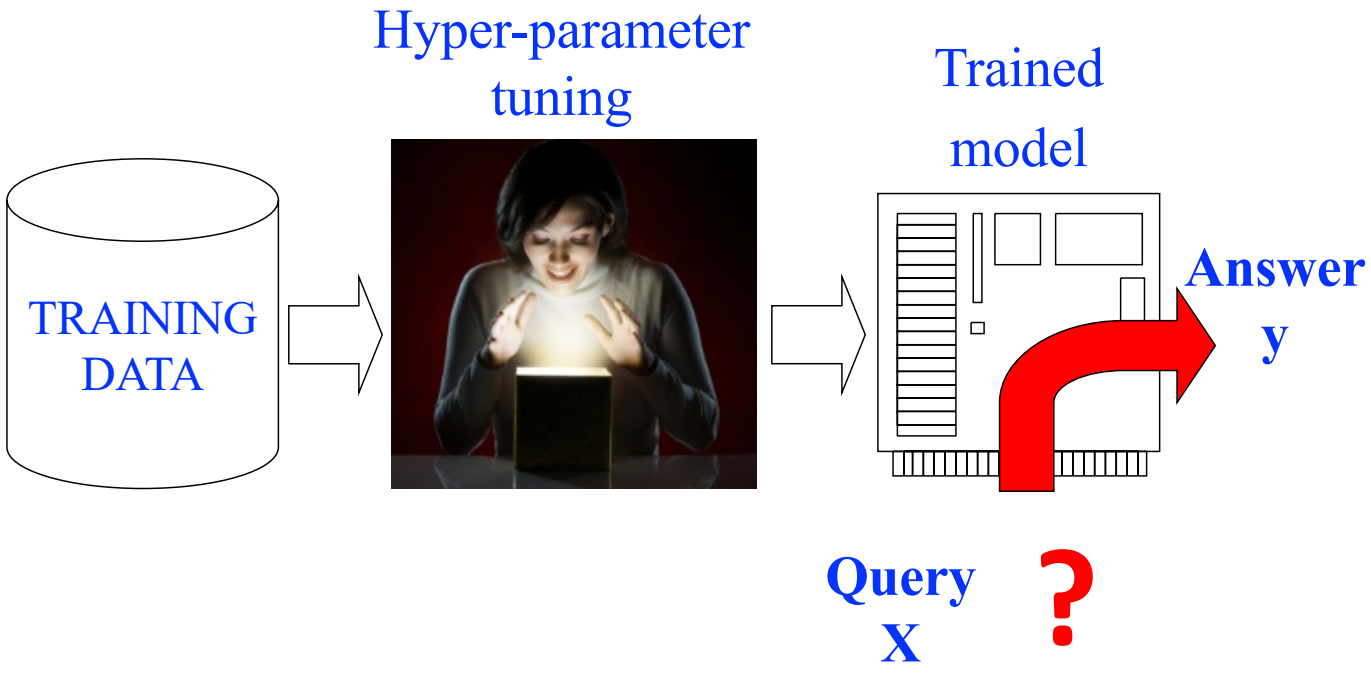
https://www.springer.com/us/book/9783030053178

* We focus on supervised learning

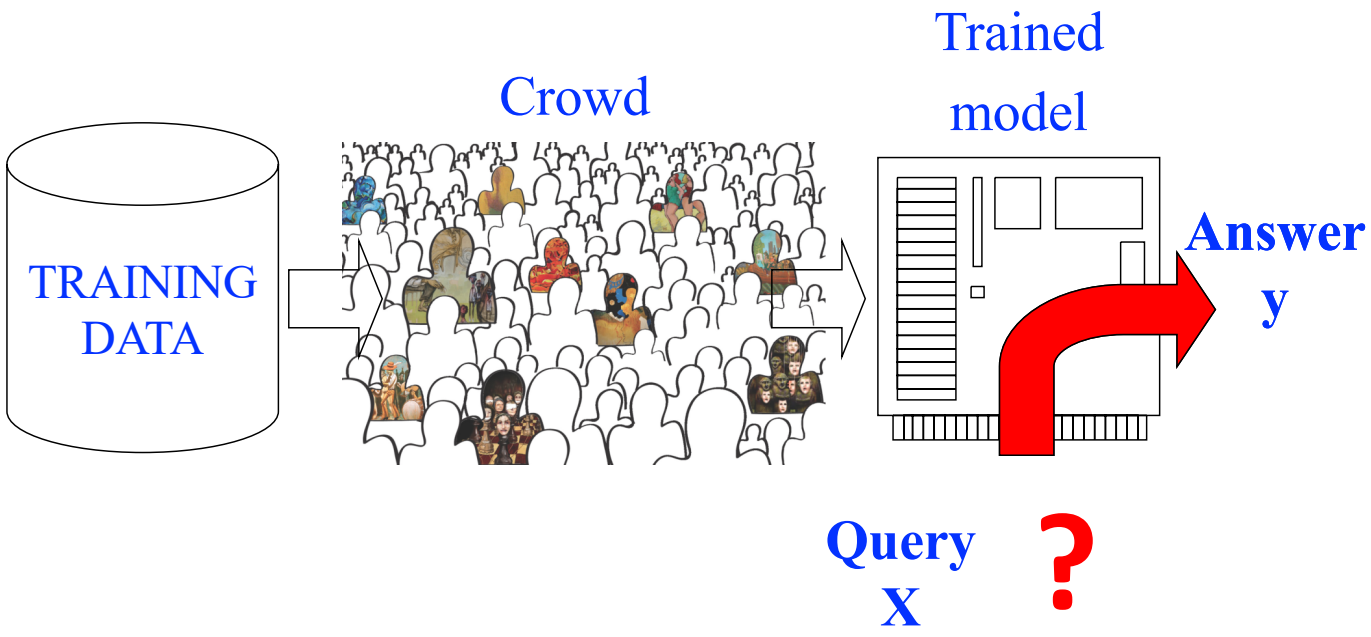
The AutoML dream



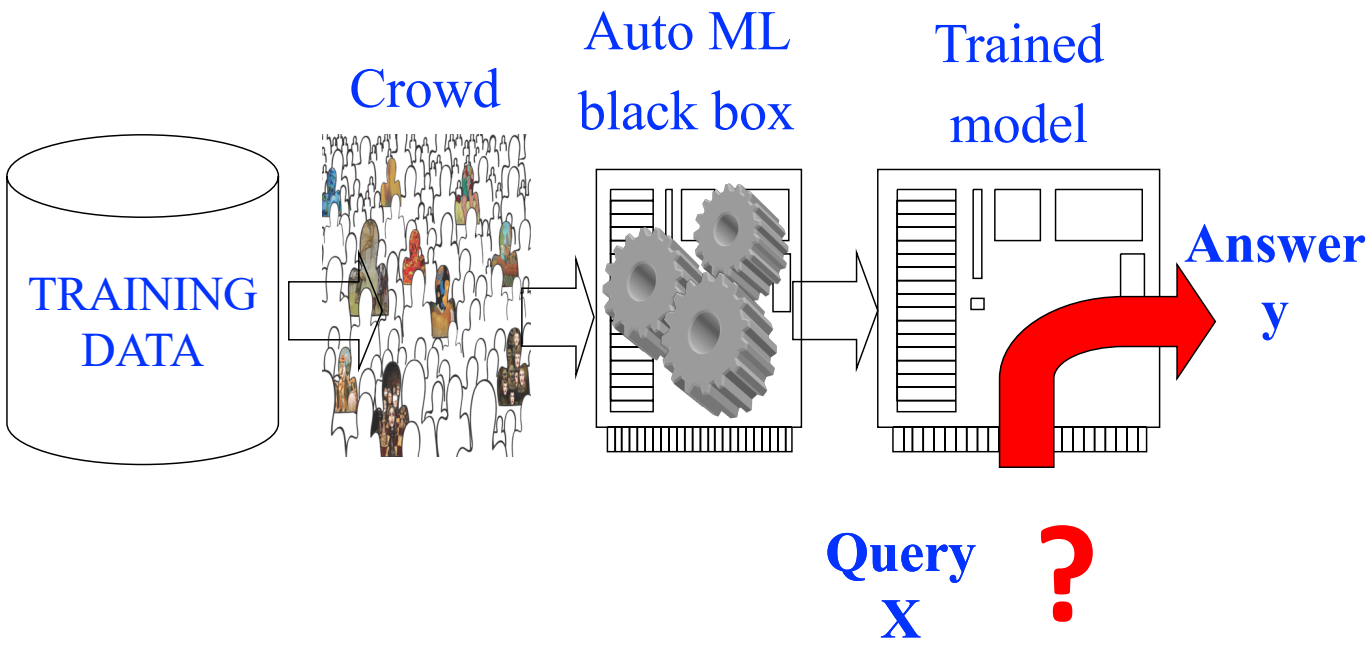
The REALITY


Hyper-parameter tuning
Trained model
TRAINING DATA
Answer y
Query X
?

ML Challenges



AutoML challenges



Brief history of AutoML challenges

- **AutoML** (2015-2016). Lasted two years divided into 5 stages. Collocated with NIPS, ICML, IJCNN. 600+ participants. 30,000USD in prizes. Sponsored by Microsoft + ChaLearn
  - Winner: AAD_Freiburg

- **AutoML2** (2018). Lasted about 4 months. Collocated with PAKDD18. 250+ participants. 10,000USD in prizes. Sponsored by 4Paradigm + ChaLearn.
  - Winner: AAD_Freiburg

- **AutoML3** (2018-2019). Lasted ~2 months. Collocated with NeurIPS18, rematches at IJCNN2019, PAKDD2019. 15,000+USD in prizes. Sponsored by 4Paradigm + Microsoft + Chalearn.
  - Winner: Autodidact.AI (ongoing!)

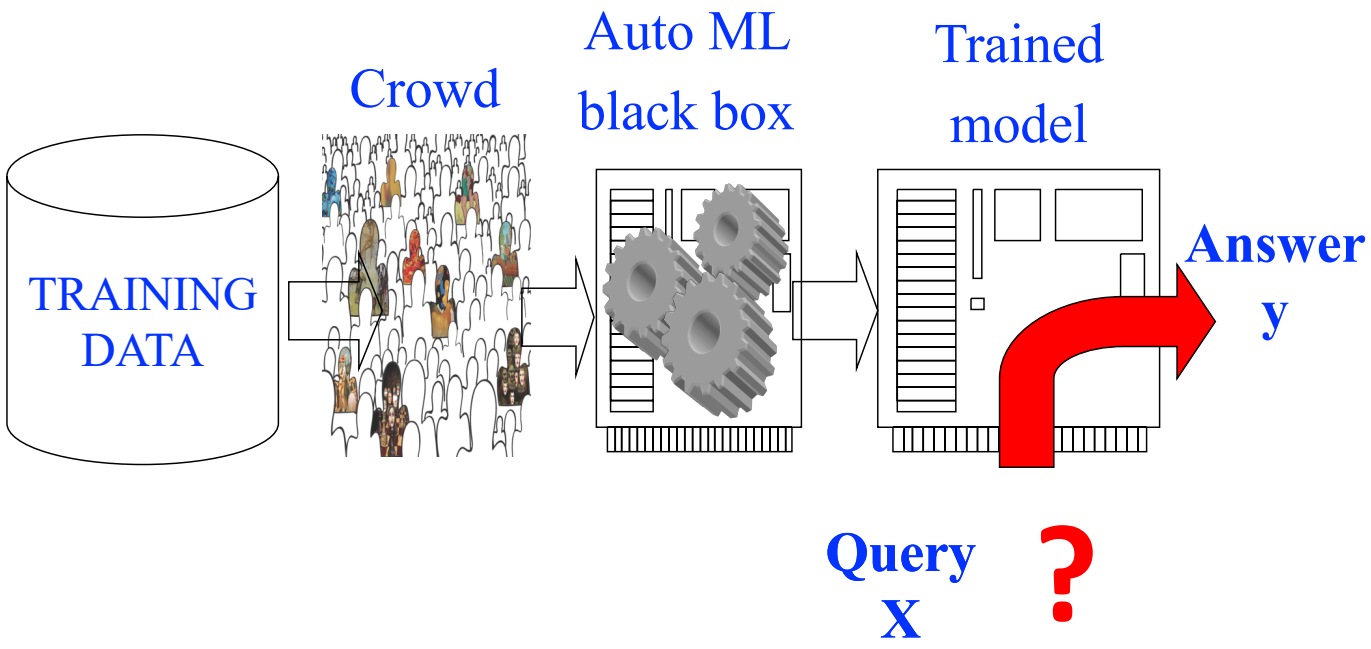
The AutoML challenge on

CodaLab ALPHA

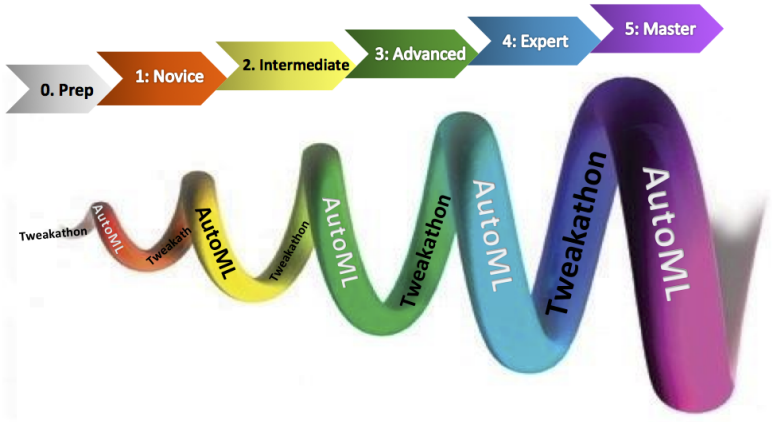
Isabelle Guyon, Kristin Bennett, Gavin Cawley, Hugo Jair Escalante, Sergio Escalera, Tin Kam Ho, Nuria Macia, Bisakha Ray, Mehreen Saeed, Alexander Statnikov, Evelyne Viegas

*automl@**ChaLearn**.org*

AutoML challenges



Rounds



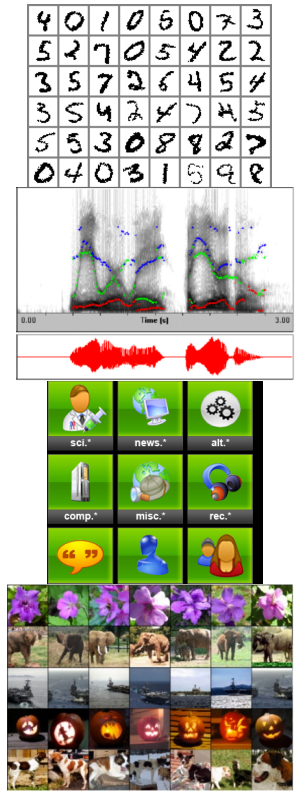
1. **NOVICE:** Binary classification.
2. **INTERMEDIATE:** Multiclass classification.
3. **ADVANCED:** Multiclass and multilabel.
4. **EXPERT:** Classification and regression.
5. **MASTER:** All of the above.

- **AutoML:** Automatic code execution on Codalab platform.
- **Tweakathon:** Result or code submission.
- To earn prizes: code should be made open source.

Data: 30 large datasets

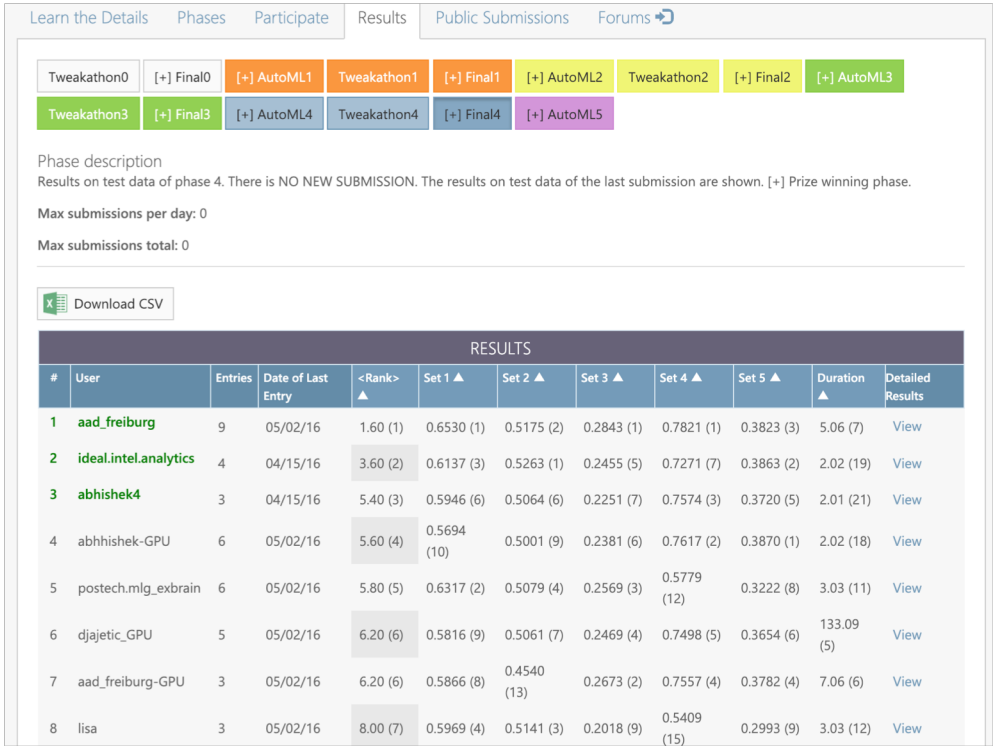
http://automl.chalearn.org/data

Round	Num	Name	Task	Metric	Time	Cnum	Cbal	Sparse	Missng	Catvar	Irrvar	Pte	Pva	Ptr	N	Ptr/N
0	1	ADULT	multilabel	F1	300	3	1	0.16	0.011	1	0.5	9768	4884	34190	24	1,424.58
0	2	CADATA	regression	R2	200	0	NaN	0	0	0	0.5	10640	5000	5000	16	312.5
0	3	DIGITS	multiclass	BAC	300	10	1	0.42	0	0	0.5	35000	20000	15000	1568	9.57
0	4	DOROTHEA	binary	AUC	100	2	0.46	0.99	0	0	0.5	800	350	800	100000	0.01
0	5	NEWSGROUPS	multiclass	PAC	300	20	1	1	0	0	0	3755	1877	13142	61188	0.21
1	1	CHRISTINE	binary	BAC	1200	2	1	0.071	0	0	0.5	2084	834	5418	1636	3.31
1	2	JASMINE	binary	BAC	1200	2	1	0.78	0	0	0.5	1756	526	2984	144	20.72
1	3	MADELINE	binary	BAC	1200	2	1	1.2 E-06	0	0	0.92	3240	1080	3140	259	12.12
1	4	PHILIPPINE	binary	BAC	1200	2	1	0.0012	0	0	0.5	4664	1166	5832	308	18.94
1	5	SYLVINE	binary	BAC	1200	2	1	0.01	0	0	0.5	10244	5124	5124	20	256.2
2	1	ALBERT	binary	F1	1200	2	1	0.049	0.14	1	0.5	51048	25526	425240	78	5,451.79
2	2	DILBERT	multiclass	PAC	1200	5	1	0	0	0	0.16	9720	4860	10000	2000	5
2	3	FABERT	multiclass	PAC	1200	7	0.96	0.99	0	0	0.5	2354	1177	8237	800	10.3
2	4	ROBERT	multiclass	BAC	1200	10	1	0.01	0	0	0	5000	2000	10000	7200	1.39
2	5	VOLKERT	multiclass	PAC	1200	10	0.89	0.34	0	0	0	7000	3500	58310	180	323.94
3	1	ALEXIS	multilabel	AUC	1200	18	0.92	0.98	0	0	0	15569	7784	54491	5000	10.9
3	2	DIONIS	multiclass	BAC	1200	355	1	0.11	0	0	0	12000	6000	416188	60	6,936.47
3	3	GRIGORIS	multilabel	AUC	1200	91	0.87	1	0	0	0	9920	6486	45400	301561	0.15
3	4	JANNIS	multiclass	BAC	1200	4	0.8	7.3 E-05	0	0	0.5	9851	4926	83733	54	1,550.61
3	5	WALLIS	multiclass	AUC	1200	11	0.91	1	0	0	0	8196	4098	10000	193731	0.05
4	1	EVITA	binary	AUC	1200	2	0.21	0.91	0	0	0.46	14000	8000	20000	3000	6.67
4	2	FLORA	regression	ABS	1200	0	NaN	0.99	0	0	0.25	2000	2000	15000	200000	0.08
4	3	HELENA	multiclass	BAC	1200	100	0.9	6 E-05	0	0	0	18628	9314	65196	27	2,414.67
4	4	TANIA	multilabel	PAC	1200	95	0.79	1	0	0	0	44635	22514	157599	47236	3.34
4	5	YOLANDA	regression	R2	1200	0	NaN	1 E-07	0	0	0.1	30000	30000	400000	100	4000
5	1	ARTURO	multiclass	F1	1200	20	1	0.82	0	0	0.5	2733	1366	9565	400	23.91
5	2	CARLO	binary	PAC	1200	2	0.097	0.0027	0	0	0.5	10000	10000	50000	1070	46.73
5	3	MARCO	multilabel	AUC	1200	180	0.76	0.99	0	0	0	20482	20482	163860	15299	10.71
5	4	PABLO	regression	ABS	1200	0	NaN	0.11	0	0	0.5	23565	23565	188524	120	1,571.03
5	5	WALDO	multiclass	BAC	1200	4	1	0.029	0	1	0.5	2430	2430	19439	270	72



Leaderboard

- Leaderboard for both tracks in every stages
  - Validation leaderboard
  - Final evaluation leaderboard



Learn the Details | Phases | Participate | Results | Public Submissions | Forums

Tweakathon0 | [+] Final0 | [+] AutoML1 | Tweakathon1 | [+] Final1 | [+] AutoML2 | Tweakathon2 | [+] Final2 | [+] AutoML3
Tweakathon3 | [+] Final3 | [+] AutoML4 | Tweakathon4 | [+] Final4 | [+] AutoML5

Phase description
Results on test data of phase 4. There is NO NEW SUBMISSION. The results on test data of the last submission are shown. [+] Prize winning phase.

Max submissions per day: 0

Max submissions total: 0

Download CSV

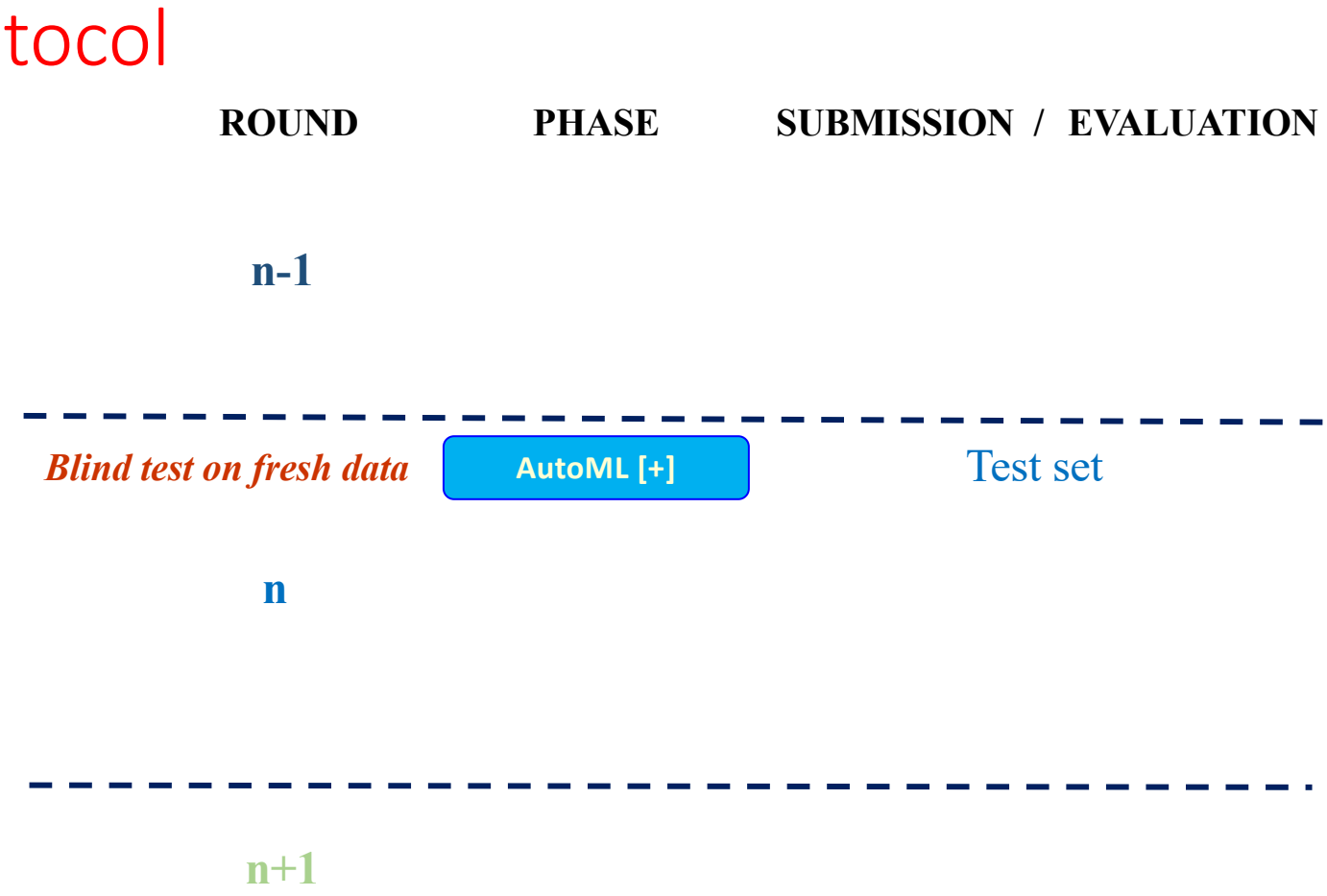
RESULTS

#	User	Entries	Date of Last Entry	<Rank> ▲	Set 1 ▲	Set 2 ▲	Set 3 ▲	Set 4 ▲	Set 5 ▲	Duration ▲	Detailed Results
1	aad_freiburg	9	05/02/16	1.60 (1)	0.6530 (1)	0.5175 (2)	0.2843 (1)	0.7821 (1)	0.3823 (3)	5.06 (7)	View
2	ideal.intel.analytics	4	04/15/16	3.60 (2)	0.6137 (3)	0.5263 (1)	0.2455 (5)	0.7271 (7)	0.3863 (2)	2.02 (19)	View
3	abhishek4	3	04/15/16	5.40 (3)	0.5946 (6)	0.5064 (6)	0.2251 (7)	0.7574 (3)	0.3720 (5)	2.01 (21)	View
4	abhhishek-GPU	6	05/02/16	5.60 (4)	0.5694 (10)	0.5001 (9)	0.2381 (6)	0.7617 (2)	0.3870 (1)	2.02 (18)	View
5	postech.mlg_exbrain	6	05/02/16	5.80 (5)	0.6317 (2)	0.5079 (4)	0.2569 (3)	0.5779 (12)	0.3222 (8)	3.03 (11)	View
6	djajetic_GPU	5	05/02/16	6.20 (6)	0.5816 (9)	0.5061 (7)	0.2469 (4)	0.7498 (5)	0.3654 (6)	133.09 (5)	View
7	aad_freiburg-GPU	3	05/02/16	6.20 (6)	0.5866 (8)	0.4540 (13)	0.2673 (2)	0.7557 (4)	0.3782 (4)	7.06 (6)	View
8	lisa	3	05/02/16	8.00 (7)	0.5969 (4)	0.5141 (3)	0.2018 (9)	0.5409 (15)	0.2993 (9)	3.03 (12)	View

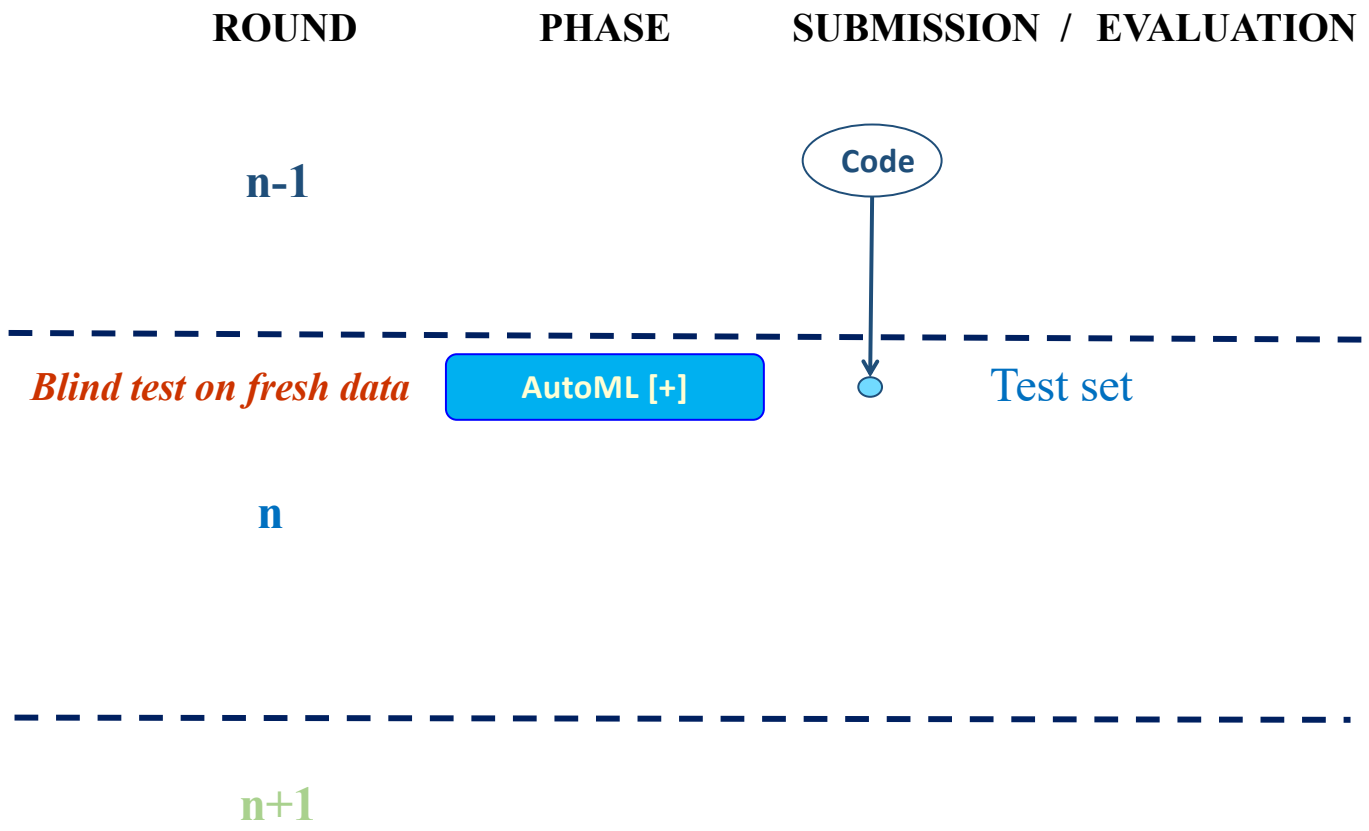
Protocol

ROUND	PHASE	SUBMISSION / EVALUATION
n-1		
n		
n+1		

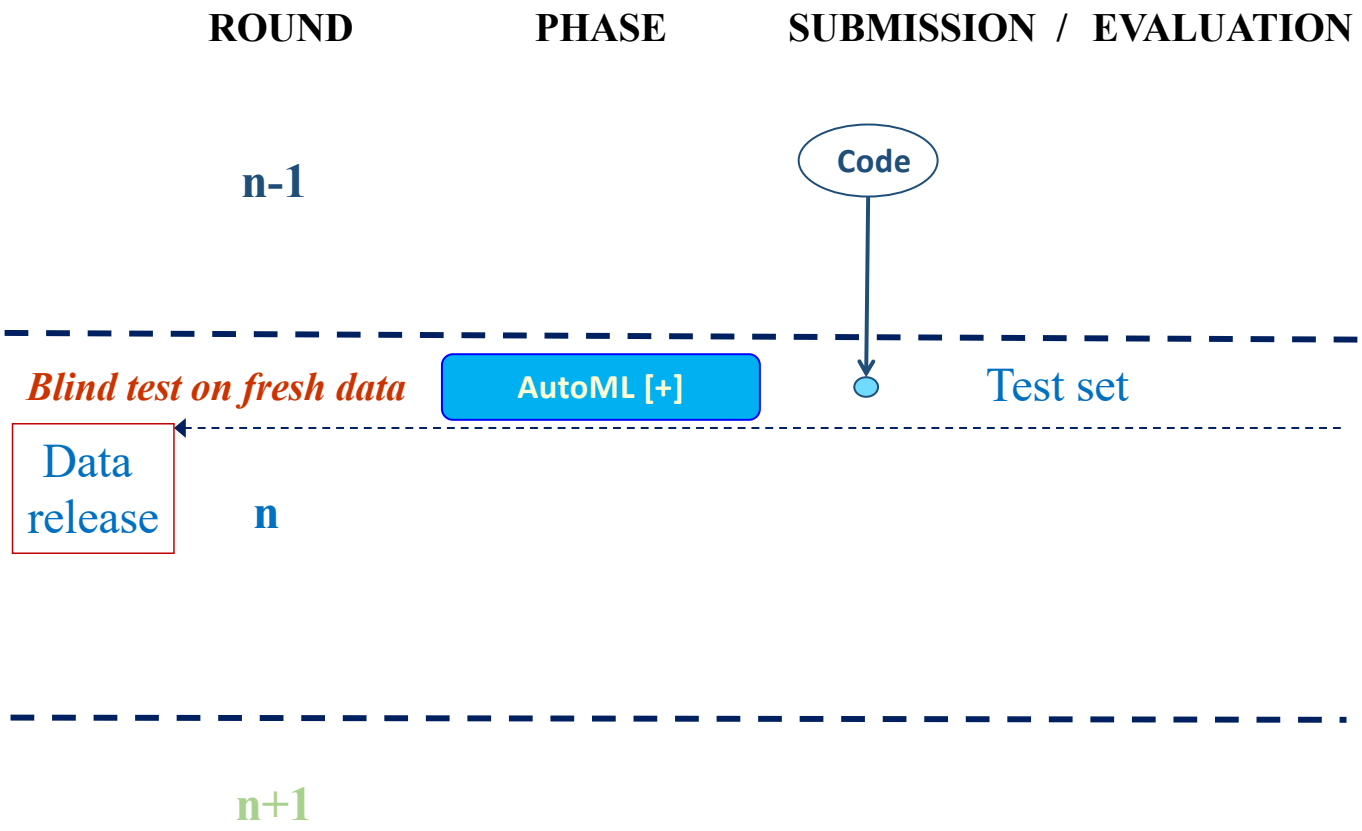
Protocol


ROUND
PHASE
SUBMISSION / EVALUATION
n-1
Blind test on fresh data
AutoML [+]
Test set
n
n+1

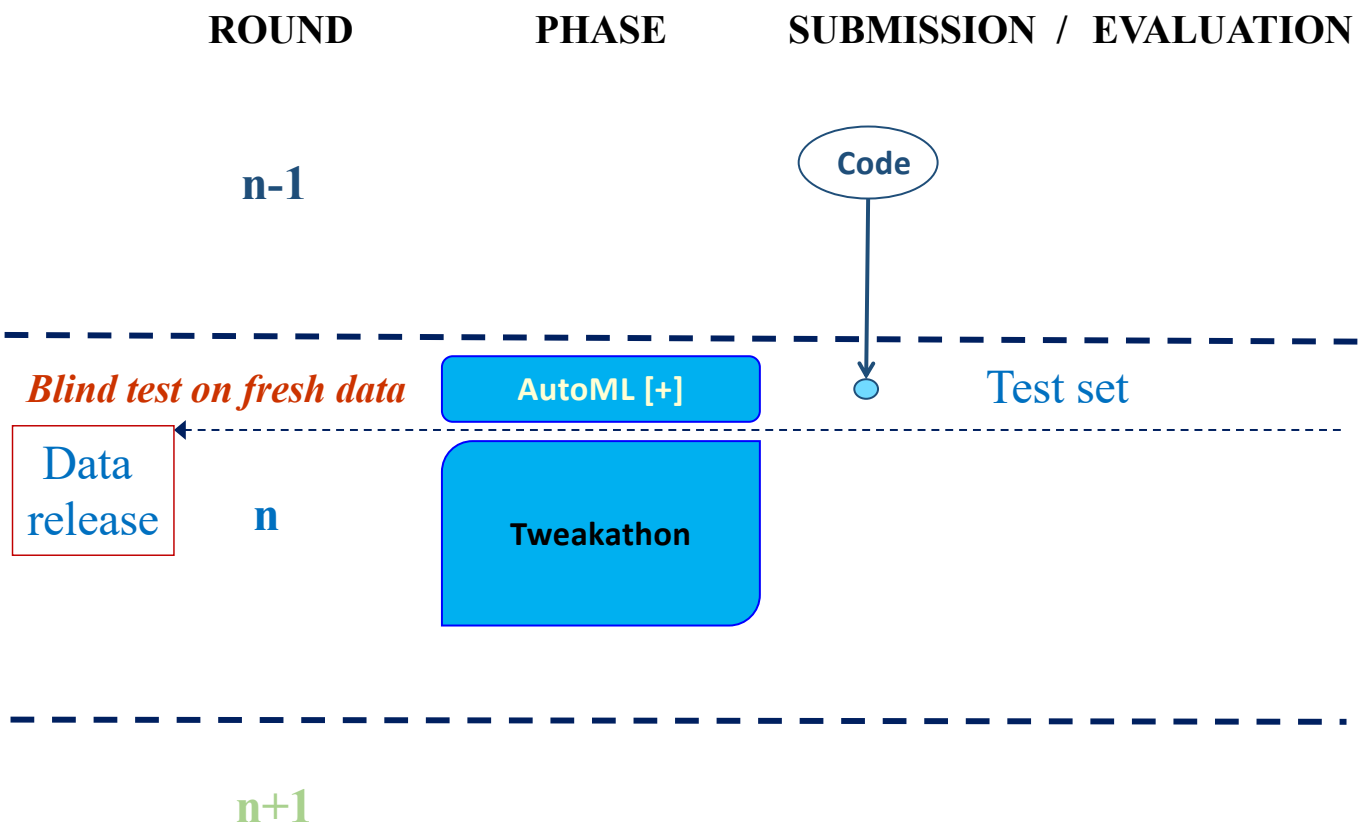
Protocol


ROUND
PHASE
SUBMISSION / EVALUATION
n-1
Code
Blind test on fresh data
AutoML [+]
Test set
n
n+1

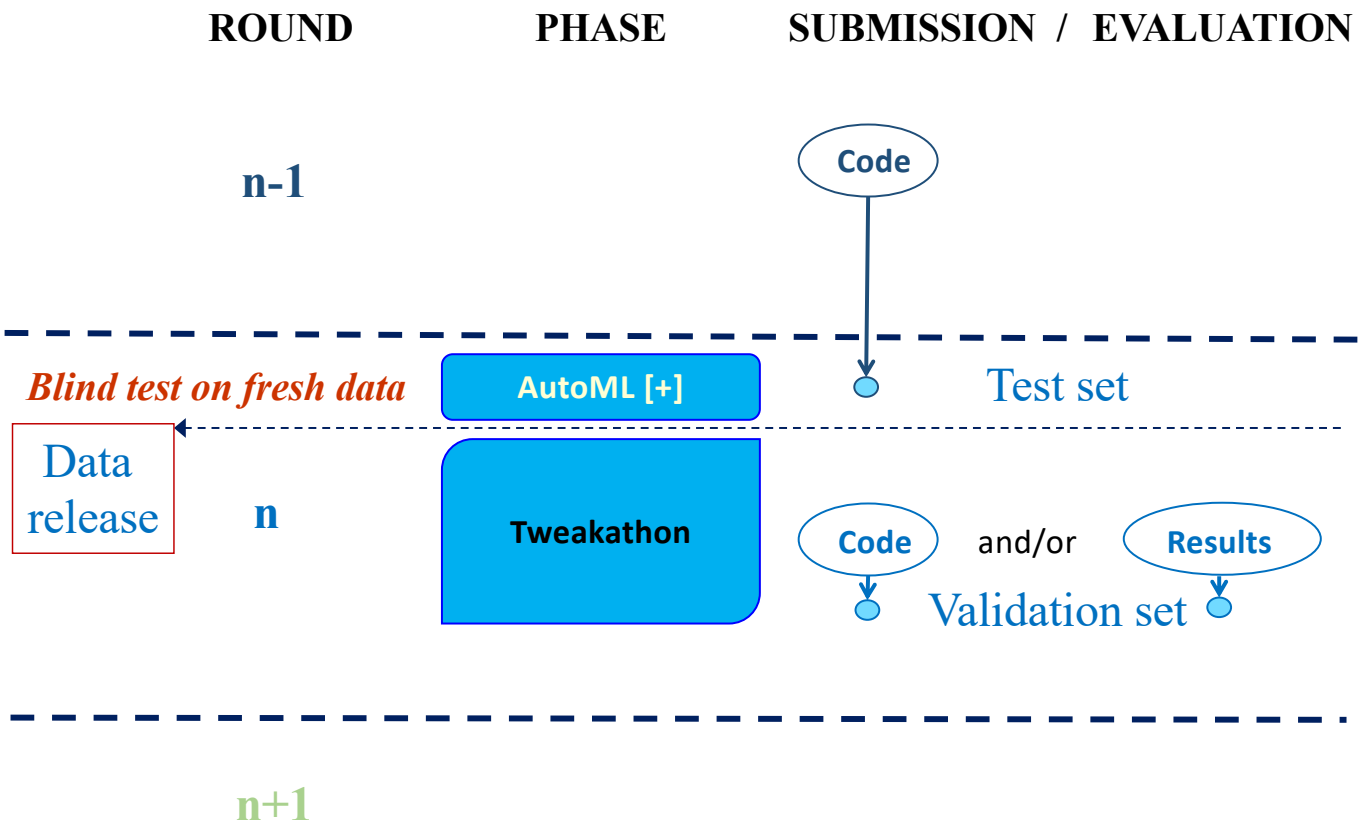
Protocol


ROUND
PHASE
SUBMISSION / EVALUATION
n-1
Code
Blind test on fresh data
AutoML [+]
Test set
Data release
n
n+1

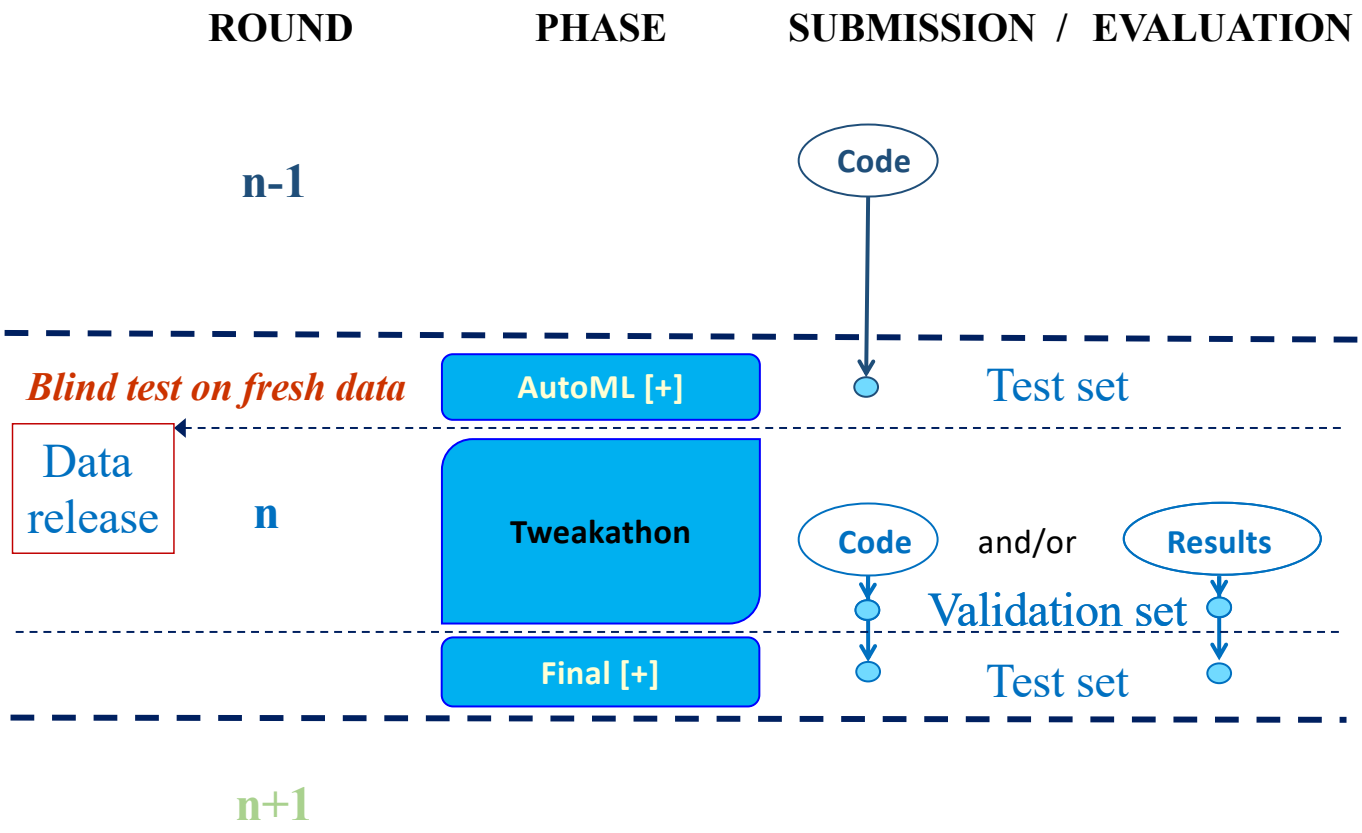
Protocol


ROUND
PHASE
SUBMISSION / EVALUATION
n-1
Code
Blind test on fresh data
AutoML [+]
Test set
Data release
n
Tweakathon
n+1

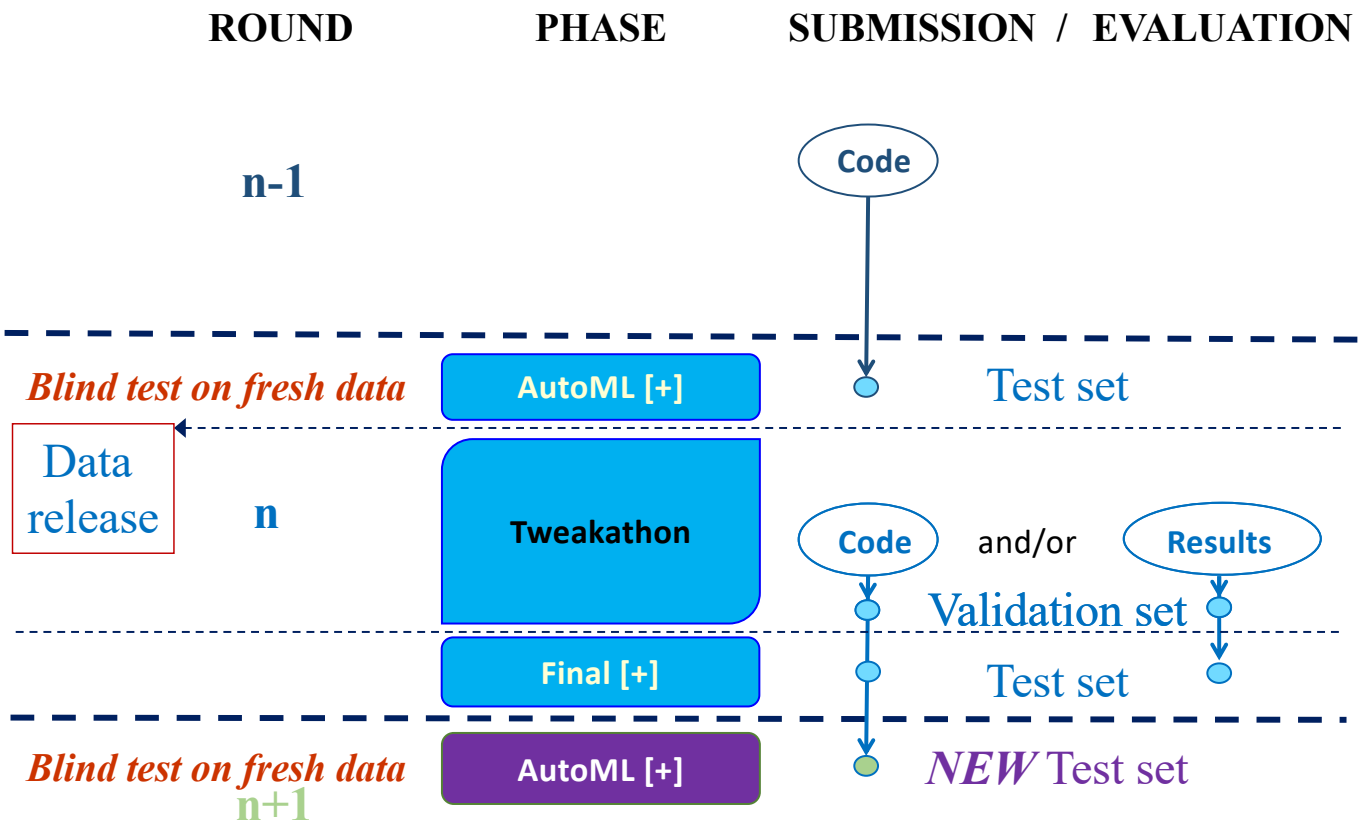
Protocol


ROUND
PHASE
SUBMISSION / EVALUATION
n-1
Code
Blind test on fresh data
AutoML [+]
Test set
Data release
n
Tweakathon
Code
and/or
Results
Validation set
n+1

Protocol


ROUND
PHASE
SUBMISSION / EVALUATION
n-1
Code
Blind test on fresh data
AutoML [+]
Test set
Data release
n
Tweakathon
Code
and/or
Results
Validation set
Final [+]
Test set
n+1

Protocol


ROUND
PHASE
SUBMISSION / EVALUATION
n-1
Code
Blind test on fresh data
AutoML [+]
Test set
Data release
n
Tweakathon
Code
and/or
Results
Validation set
Final [+]
Test set
Blind test on fresh data
n+1
AutoML [+]
NEW Test set

Best overall AutoML: aad_freiburg

Rnd	AutoML Ended	AutoML Winners	AutoML $< R >$	AutoML $< S >$	Final Ended	Final Winners	Final $< R >$	Final $< S >$	UP (%)
0	NA	NA	NA	NA	02/14/15	1. ideal 2. abhi 3. aad	1.40 3.60 4.00	0.8159 0.7764 0.7714	NA
1	02/15/15	1. aad 2. jrl44 3. tadej	2.80 3.80 4.20	0.6401 0.6226 0.6456	06/14/15	1. aad 2. ideal 3. amsl	2.20 3.20 4.60	0.7479 0.7324 0.7158	15
2	06/15/15	1. jrl44 2. aad 3. mat	1.80 3.40 4.40	0.4320 0.3529 0.3449	11/14/15	1. ideal 2. djaj 3. aad	2.00 2.20 3.20	0.5180 0.5142 0.4977	35
3	11/15/15	1. djaj 2. NA 3. NA	2.40 NA NA	0.0901 NA NA	02/19/16	1. aad 2. djaj 3. ideal	1.80 2.00 3.80	0.8071 0.7912 0.7547	481
4	02/20/16	1. aad 2. djaj 3. marc	2.20 2.20 2.60	0.3881 0.3841 0.3815	05/1/16	1. aad 2. ideal 3. abhi	1.60 3.60 5.40	0.5238 0.4998 0.4911	31
GPU	NA	NA	NA	NA	05/1/16	1. abhi 2. djaj 3. aad	5.60 6.20 6.20	0.4913 0.4900 0.4884	NA
5	05/1/16	1. aad 2. djaj 3. post	1.60 2.60 4.60	0.5282 0.5379 0.4150	NA	NA	NA	NA	NA

aad=aad_freiburg djaj=djajetic marc=marc.boulle tadej=tadejs
abhi=abhishek4 ideal=ideal.intel.analytics mat=matthias.vonrohr
asml=amsl.intel.com jlr44 = backstreet.bayes post = postech.mlg_exbrain

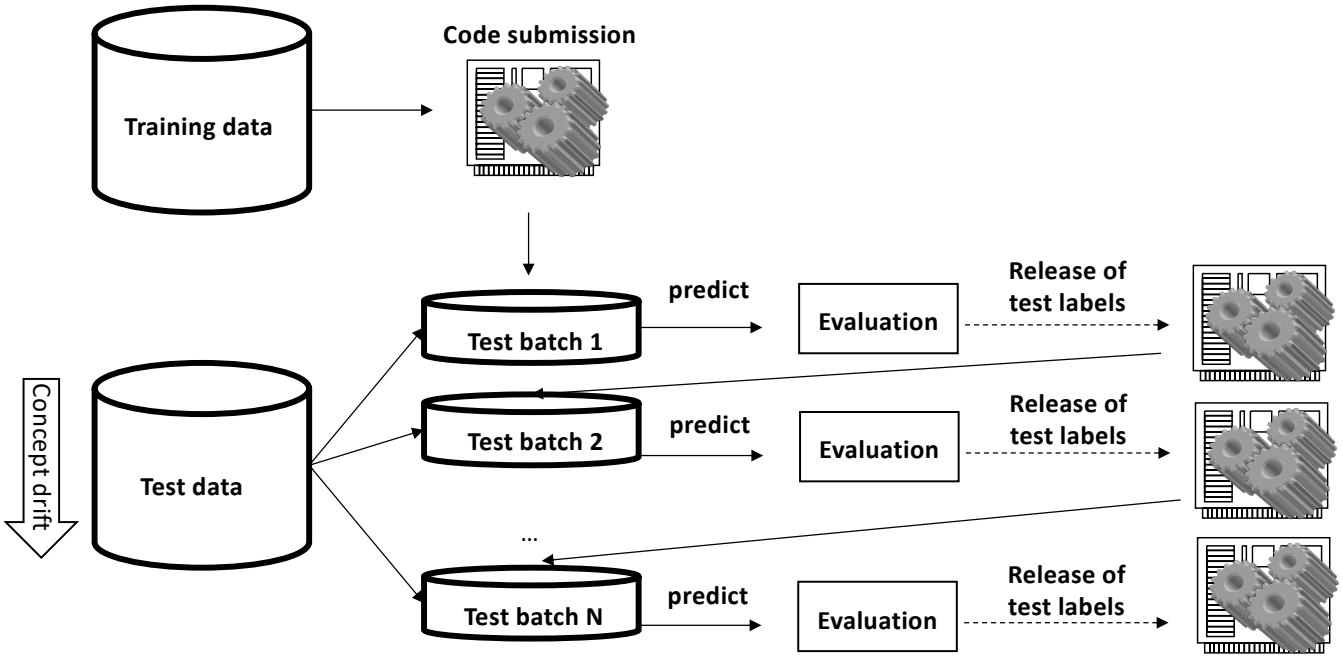
AutoML 2018-2019 challenges

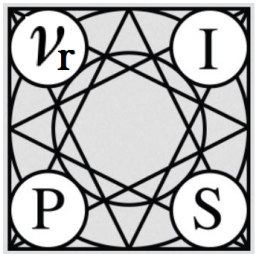
- Two other editions have been organized since then.
- Only code submission was considered!

AutoML 2018-2019 challenges

- AutoML has been challenged by:
  - **Extreme imbalance ratios.** ~1% imbalance
  - **Scalability.** Data sets larger than ever will be considered (Up to 10M instances)
  - **Concept drift.** Dependency between instances, concept changing through time.
  - **LifeLong setting.** Evaluation of the lifelong capabilities of learning machines.
  - **Mixed features.** Including numerical, categorical, time based, and Multi-value Categorical features.

AutoML3@NeurIPS2018





AutoML for Lifelong Machine Learning

Hugo Jair Escalante, Wei-Wei Tu, Isabelle Guyon, Daniel Silver,
Evelyne Viegas, Yuqiang Chen, Qiang Yang

- Duration: ~3 months
- Participants: 334
- Outcomes:
  - Baseline outperformed by a large margin
  - Novel, original solutions
  - Data set size was the main challenge
  - Several top ranked participants failed to fit their solutions to the available resources

1st place. Autodidact.ai. *Jobin Wilson, Amit Kumar Meher, Bivin Vinodkumar Bindu, Manoj Sharma, Vishakha Pareek.* Flytxt, Indian Institute of Technology Delhi, CSIR-CEERI

2nd place. Meta_Learners. *Zheng Xiong, Jiyan Jiang, Wenpeng Zhang* Tsinghua University, China

3rd place. GrandMasters. *Jiangeng Chang, Yakun Zhao, Honggang Liu, Jinlong Chai.* BeiJing University of Post and Telecom WCSN Lab, BeiJing University of Post and Telecom AI & HPC Department. Inspur Electronic Central South University, China

Organizers

Microsoft

université
PARIS-SACLAY

Brief history of AutoML challenges

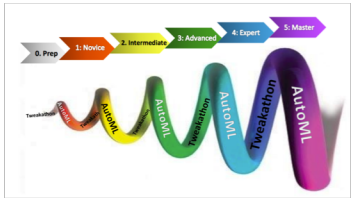
- **AutoML** (2015-2016). Lasted two years divided into 5 stages. Collocated with NIPS, ICML, IJCNN. 600+ participants. 30,000USD in prizes. Sponsored by Microsoft + ChaLearn
  - Winner: AAD_Freiburg

- **AutoML2** (2018). Lasted about 4 months. Collocated with PAKDD18. 250+ participants. 10,000USD in prizes. Sponsored by 4Paradigm + ChaLearn.
  - Winner: AAD_Freiburg

- **AutoML3** (2018-2019). Lasted ~2 months. Collocated with NeurIPS18, rematches at IJCNN2019, PAKDD2019. 15,000+USD in prizes. Sponsored by 4Paradigm + Microsoft + Chalearn.
  - Winner: Autodidact.AI (ongoing!)

AutoML challenges 2015 – 2018

- Advancing the SOTA in AutoML
  - Solutions publicly available: AutoSKLearn
- A few steps towards reproducibility of results in challenges
  - Among the first ML challenges to support code submission
  - Comparison of offline and online submissions
  - All challenge participants subject to same restrictions in terms of resources
  - Hosted among the largest research challenge with code submission
- A series of challenges : moving towards life long AutoML and adversarial AutoML.



Challenges and reproducibility

- Associating challenges to research papers!
- Key for succeeding in challenges: evaluation protocol, rules
- Making challenge organization worth for researchers

Thanks*

Hackathon team:
Marc Boullé
Lukasz Romaszco
Sébastian Treger
Emilia Vaajoensuu
Philippe Vandermersch

Software development:
Eric Carmichael
Ivan Judson
Christophe Poulain
Percy Liang
Arthur Pesah
Xavier Baro Solé
Lukasz Romaszco
Michael Zyskowski

Advisors and beta testers:
Kristin Bennett
Marc Boullé
Cecile Germain
Cecile Capponi
Richard Caruana
Gavin Cawley
Gideon Dror
Sergio Escalera
Tin Kam Ho
Balasz Kégl
Hugo Larochelle
Víctor Ponce López
Nuria Macia
Simon Mercer
Florin Popescu
Michèle Sebag
Danny Silver,

Beat AutoSKlearn
Matthias Feurer
Katharina Eggensperger
Syed Mohsin Ali
Frank Hutter

Codalab management:
Evelyne Viegas
Percy Liang
Erick Watson

AutoML Book
Frank Hutter
Roman Garnett
Joaquin Vanschoren
Lars Kotthoff

ChaLearn board/organizers

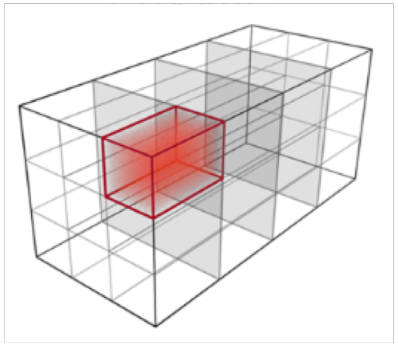
Data providers:
Yindalon Aphinyanaphongs
Olivier Chapelle
Hugo Jair Escalante
Sergio Escalera
Zainab Iftikhar Malhi
Vincent Lemaire
Chih Jen Lin
Meysam Madani
Bisakha Ray
Mehreen Saeed
Alexander Statnikov
Gustavo Stolovitzky
H-J. Thiesen
Ioannis Tsamardinos
Wei-Wei Tu
Yang Yu
Yuqiang Chen,
Qiang Yang

Result analysis:
Imad Chaabane
Lisheng Sun

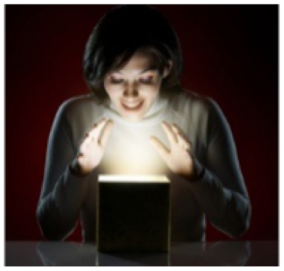
* Our apologies for missinng names

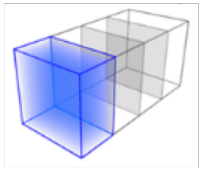
What is comming on AutoML?

Automating Deep Learning

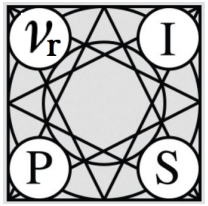


Hyper-parameter tuning





https://autodl.chalearn.org/



Call for NeurIPS 2019 competitions

We invite **proposals for the 2018 Neural Information Processing Systems Competition track** (NIPS 2018: https://nips.cc/Conferences/2018) in Montréal, Canada. After the success of the first NIPS 2017 Competition track, a second edition of the Competitions track will be held at NIPS 2018. We solicit competition proposals on any topic of interests to the NIPS community. We especially encourage competition proposals from emerging new fields or new application domains related to NIPS. Interdisciplinary topics that could attract a significant cross-section of the community are highly valued.

There will be two kinds of competitions:

1. Standard **data science driven competitions**, where participants will compete to obtain the best score on a machine learning problem of interest to the NIPS community based on a problem and data defined and released by the organizers of the competition.
2. **Live competitions**, which will be held in a science-fair manner at NIPS. Participants will present live demos at NIPS which apply methodology in an application domain defined by the organizers of the Live competition.

There will be a Competition track session on December 7 where competition results can be discussed and presented. Organizers will propose a tentative schedule for the presentation of the competition and its results based on the assigned time slot. The main conference will provide coffee breaks and, if necessary, poster facilities. For any additional questions please contact the competition chairs.

- Competition chairs:
  - Hugo Jair Escalante, INAOE, ChaLearn, hugo.jair@gmail.com
  - Ralf Herbrich, Amazon, herbrich@amazon.com

Forthcoming events/activities

- Special session on AutoML @ IJCNN2019

- IEEE TPAMI SI on Image and Video Inpainting and Denoising (**Submission deadline: December 15, 2018**)
  - http://chalearnlap.cvc.uab.es/special-issue/30/description/

- IJCAI Workshop on AutoML (TBC)

- IJCV SI on Analyzing Human Behavior from Social Media Data (**Submission deadline: March 1, 2019**)
  - http://chalearnlap.cvc.uab.es/special-issue/31/description/

- JAIR SI: AutoML (TBC: **Submission deadline: March 1, 2019**)
  - http://chalearnlap.cvc.uab.es/special-issue/31/description/

Join us for fun!

http://automl.chalearn.org/

@ChaLearn_org

automl2018@gmail.com

hugo.jair@gmail.com