

Big Data Science in Astroparticle Research: sharing and exploiting data & knowledge

Martin Erdmann & Andreas Haungs

20-Feb-2019

Workshop Program

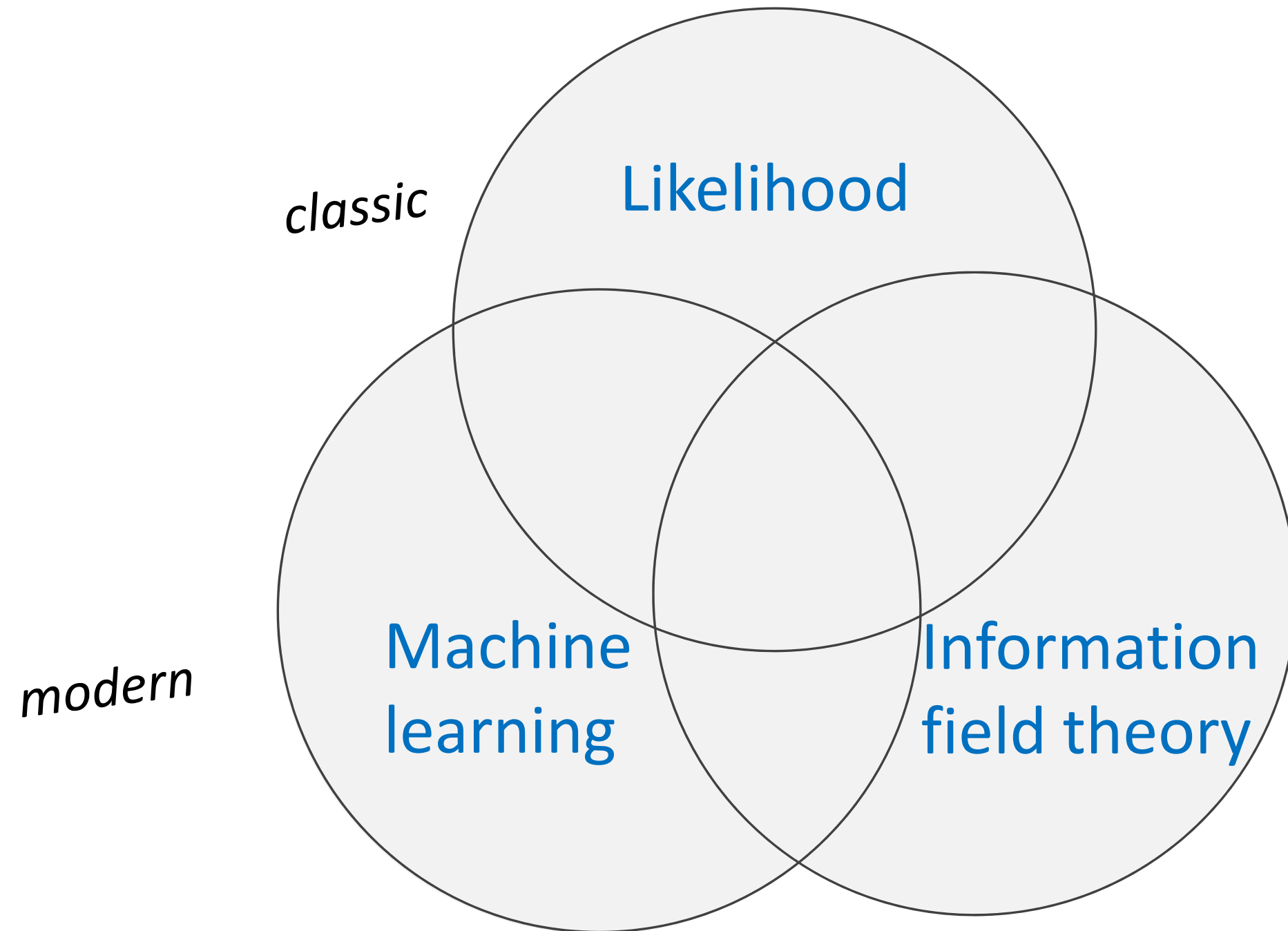
	Date	Title	Presenter
Machine Learning	Monday	14:00 Welcome and organizational matters	Prof. Martin Erdmann
	Monday	14:10 Introduction tutorial to Information Field Theory and Exercises	Prof. Torsten Enßlin
	Monday	16:15 Introduction tutorial to Information Field Theory and Exercises	Prof. Torsten Enßlin
	Monday	16:15 Beginners Tutorial Deep Networks: Introduction	Mr. Jonas Glombitza
	Tuesday	09:00 Welcome at the RWTH Aachen University	Prof. Dr. Carsten Honerkamp
	Tuesday	09:19 Deep Neural Networks for Energy and Position Reconstruction in EXO-200	Mr. Tobias Ziegler
	Tuesday	09:38 A Deep Neural Network for Pixel-Level Electromagnetic Particle Identification in the MicroBooNE Lar TPC	Mr. Davide Porzio
	Tuesday	09:57 Deep learning techniques applied to the physics of extensive air showers	Mr. Juan Miguel Carceller
	Tuesday	10:16 Inspection of the AIXNET network for air shower reconstruction	Mr. Niclas Eich
	Tuesday	11:05 Application of Deep Learning methods to analysis of Imaging Atmospheric Cherenkov Telescopes data	Mr. Matthias Büchele
	Tuesday	11:24 Recovery of Radio Signals from Cosmic Ray Induced Air Showers with Deep Learning	Mr. Felix Schlüter
	Tuesday	11:43 Reconstruction of air-shower radio pulses from real Tunka-Rex background with autoencoder network	Mr. Pavel Bezyazeekov
	Tuesday	12:02 Event classification in Compton-Pair telescopes using Deep Learning Techniques	Mr. Jan Peter Lommler
	Tuesday	12:21 Cascade Reconstruction in IceCube using Generative Networks	Mr. Mirco Muennefeld
	Tuesday	12:40 Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Net	Mr. Thorben Quast
	Tuesday	14:30 Deep Learning and Mathematical Modeling: Taking the Best out of Both Worlds	Prof. Gitta Kutyniok
	Tuesday	15:10 Contracting alignment patterns of mixed composition nuclei	Mr. Marcus Wirtz
	Tuesday	15:30 Exploring Optical Properties of Antarctic Ice with IceCube Using Gradient Descent	Mr. Alexander Harnisch
	Tuesday	16:00 Evaluation	
	Tuesday	16:20 Data Science in Astroparticle Physics	Dr. Tim Ruhe
Open Data	Tuesday	17:00 Bayesian reconstruction of the cosmic Dark Matter velocity field	Dr. Florian Führer
	Tuesday	17:20 Autonomous Physics-Inspired Feature Engineering	Mr. Marcel Rieger
	Tuesday	17:40 Exploring Antarctic Ice Properties Using Generative Neural Networks	Mr. Sebastian Bange
	Tuesday	19-22 Workshop Dinner Pontgarten	
	Wednesday	09:00 The ESCAPE project and the EOSC	Dr. Kay Graf
	Wednesday	09:40 CTA Software and Science Data Management	Dr. Igor Oya
	Wednesday	10:20 Open high-level data and tools for gamma-ray astronomy	Dr. Christoph Deil
	Wednesday	11:10 Data engineering for joint analysis of different astroparticle data in KRAD project	Mrs. Victoria Tokareva
	Wednesday	11:30 Challenges and Opportunities of Digital Transformation in Fundamental Research on Universe and Matter	Dr. Andreas Haungs
	Wednesday	12:00 Discussion on Big Data Science in Astroparticle Research	Erdmann, Haungs

This workshop: GREAT PROGRESS in 2018/19

Deep learning in astroparticle physics matured to Publications

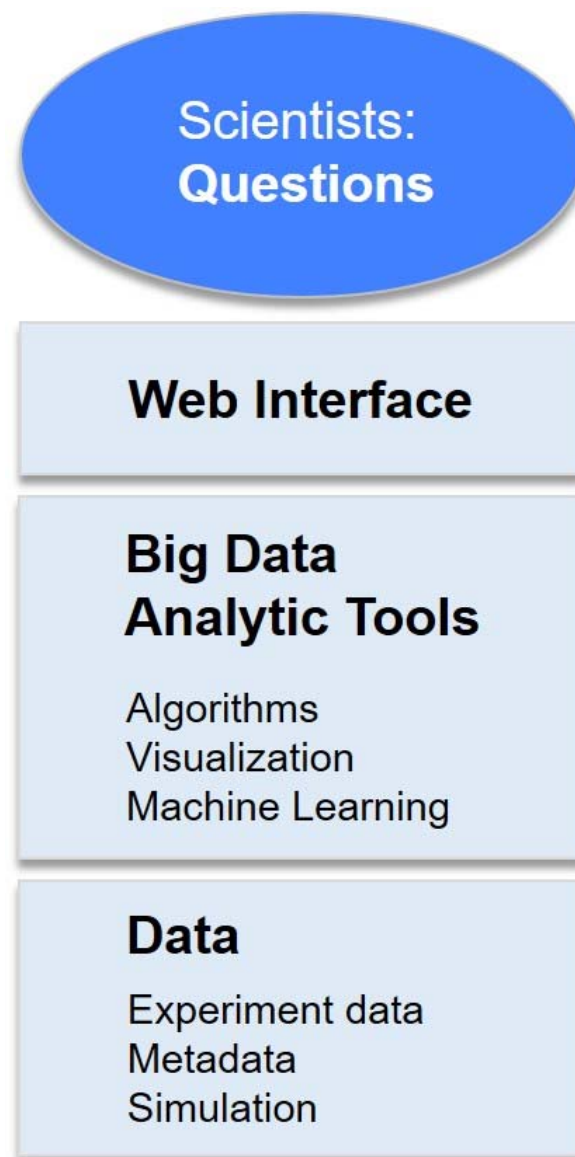
- Variants of machine learning *Mathematicians get involved in broad efforts*
- Big Data concepts *We're not alone! Senior colleagues and politicians are looking for ways to support our scientific progress.*

Select versus Combine



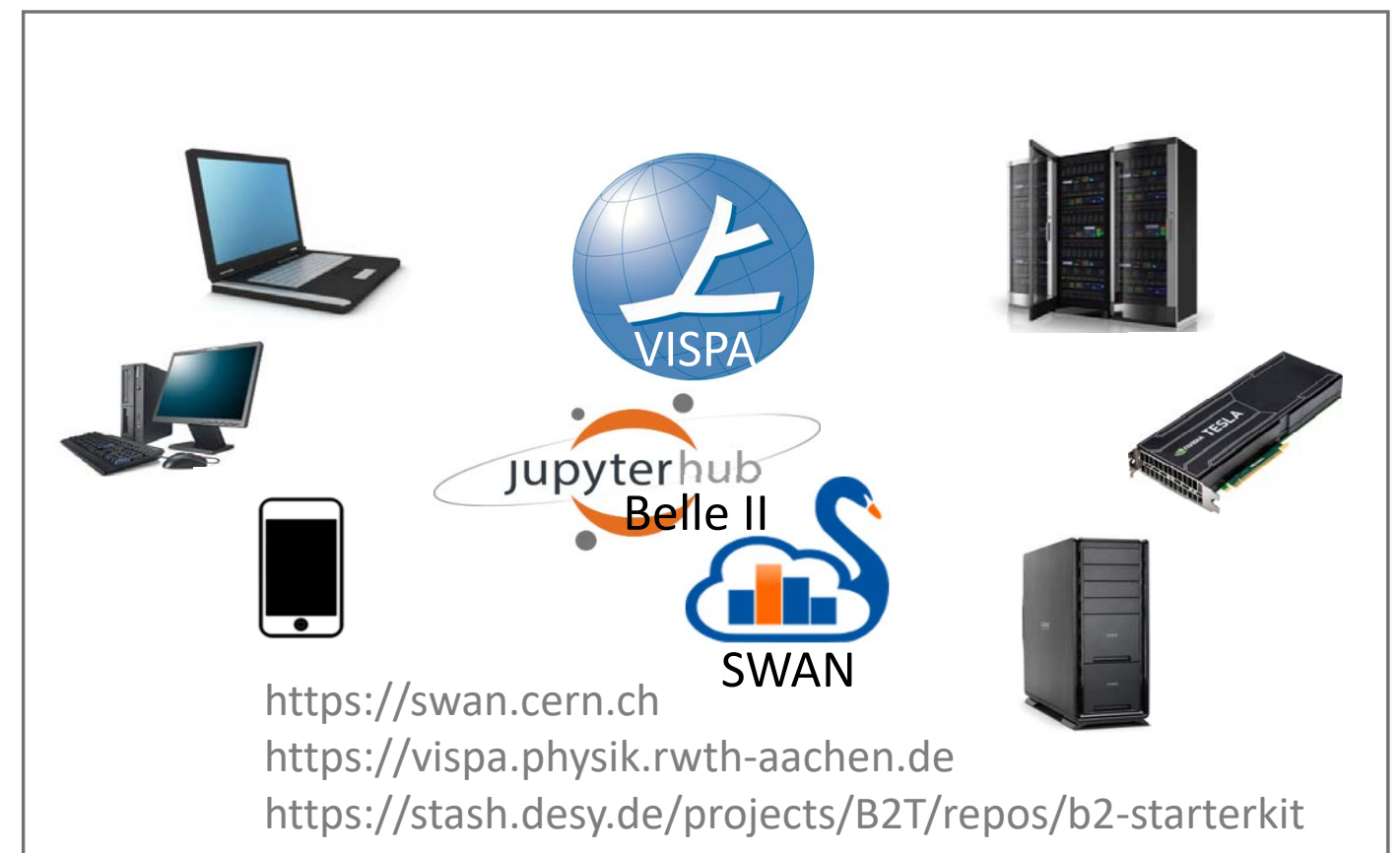
Future scientific work: EOSC, ESCAPE, SCOOGLE...

Visions are converging,
still differences in what
is leading importance



*Algorithms are the language with which
physicists form their scientific questions*

*Physics analysis workflow
e.g. Luigi Analysis Workflow*





Martin Erdmann, Andreas Haungs

Deutsche Physikalische Gesellschaft DPG: New working group „Physics, modern Information Technologie and Artificial Intelligenz“

1. BIG DATA: archiving, processing, management, analysis and simulation of complex data streams, HPC, information theory, statistical methods
2. IT: High-performance Data Readout Systems and Mass Storage, Visualization, Smart Sensors, Bridge Technologies for the Next Level of Big Data
3. AI & ROBOTICS: Data driven algorithms & software, autonomous devices, remote control, innovative applications, algorithms for quantum computers
4. HIGHER EDUCATION: Curricula and multi-disciplinary research centers, cooperation with the GI Data Scientist Task Force, IT Infrastructure
5. INDUSTRY and SOCIETY: ethics, technology assessment, sustainability, economics, law, start-ups, public relations



DPG conference Aachen 25.-29.3.2019
Arbeitskreis Physik, IT & KI (AKPIK)
Mitgliederversammlung des Arbeitskreises

16.-18.-Sep-2019

Workshop for Reconstruction and Machine Learning in Neutrino Experiments



Topics

- Reconstruction of 3D topological information
- Event classification using neural networks
- Likelihood Identification
- Projection Matching Algorithm
- and many more



At: Germany, Hamburg
DESY, building 1b, seminar room 4b



Start: Sep 16, 2019
End: Sep 18, 2019

Organisation:
Caren Hagner
Björn Wonsak

indico.desy.de/indico/event/21853/

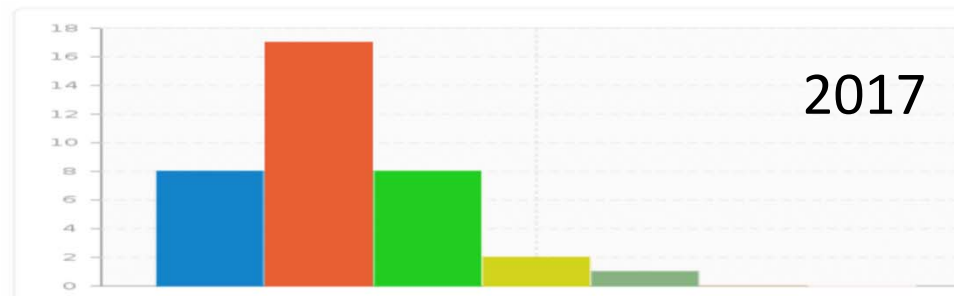
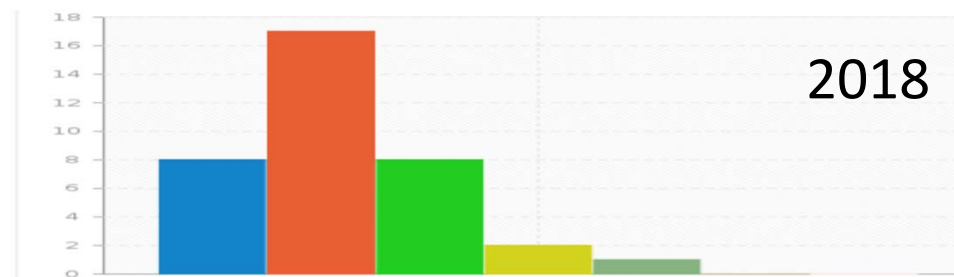
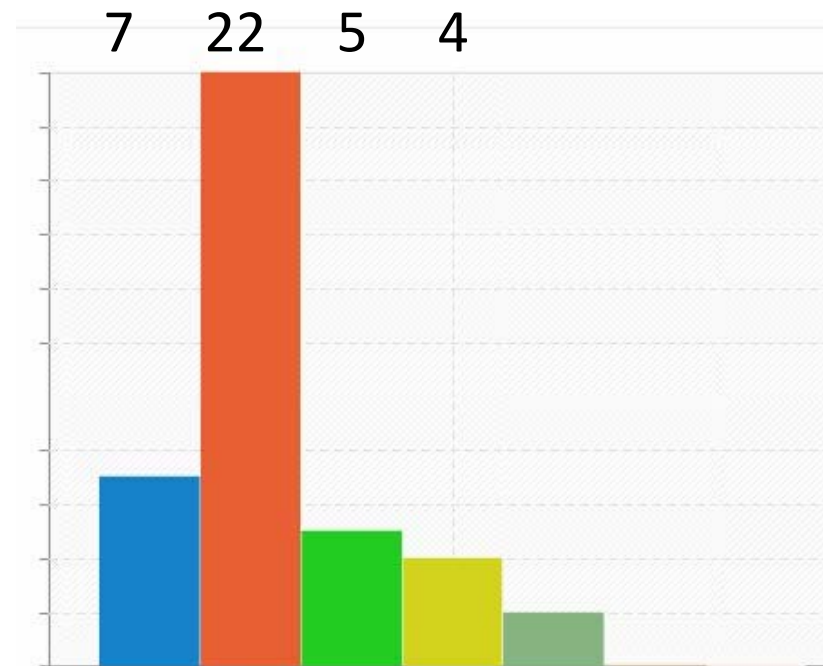


Participants

I participate in this workshop as

- Student
- PhD student
- Postdoctoral researcher
- Senior physicist (>5 years PhD)
- Other

20-Feb-2019, 40 answers of 90 participants

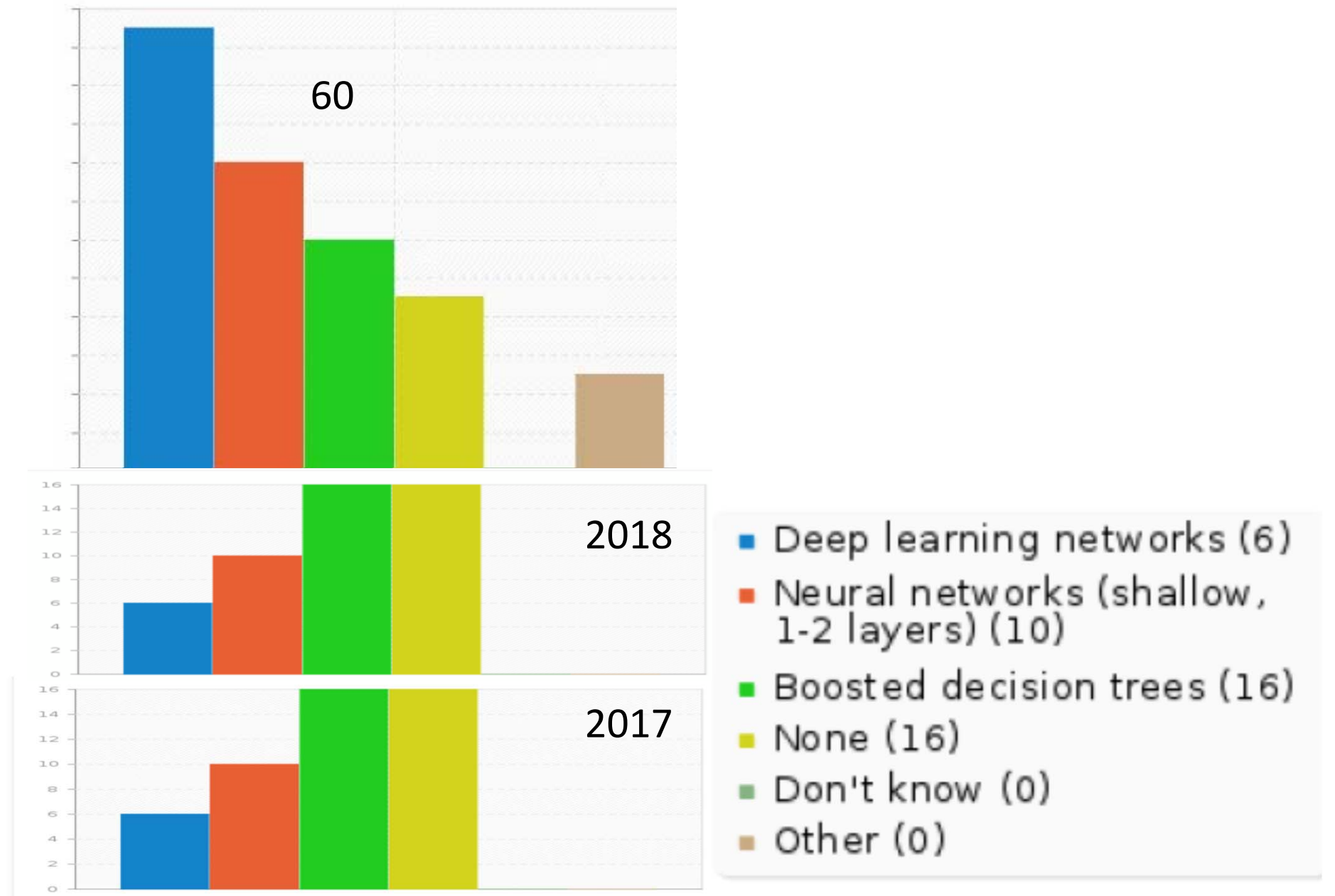


- Student (8)
- PhD student (17)
- Postdoctoral researcher (8)
- Senior physicist (>5 years PhD) (2)
- Other (1)
- No answer (0)
- Not displayed (0)

University: courses, seminars, research

I have used machine learning techniques before

- Deep learning networks
- Neural networks (shallow)
- Boosted decision trees
- None
- Don't know
- Other

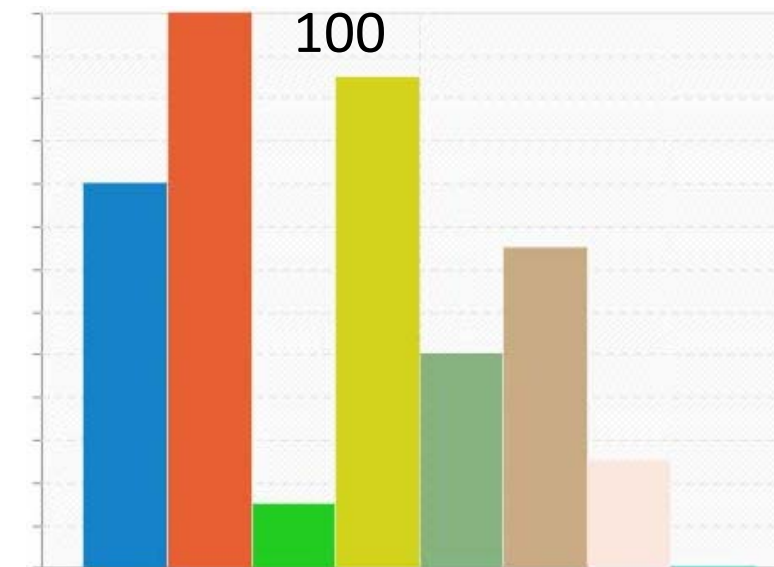


Broadcast new possibilities arising from deep learning to physics institutes

Physicists take advantage of new opportunities arising with deep networks

My research interests concerning deep learning networks are

- Image pattern recognition (e.g. event reconstruction)
- Classification (e.g. particle identification, physical processes)
- Assignments (e.g. solving ambiguities with >2 identical particles)
- Regression (calculation of variables, e.g. energy corrections)
- Adversarial training
- Curiosity
- Applications not obvious
- Other applications

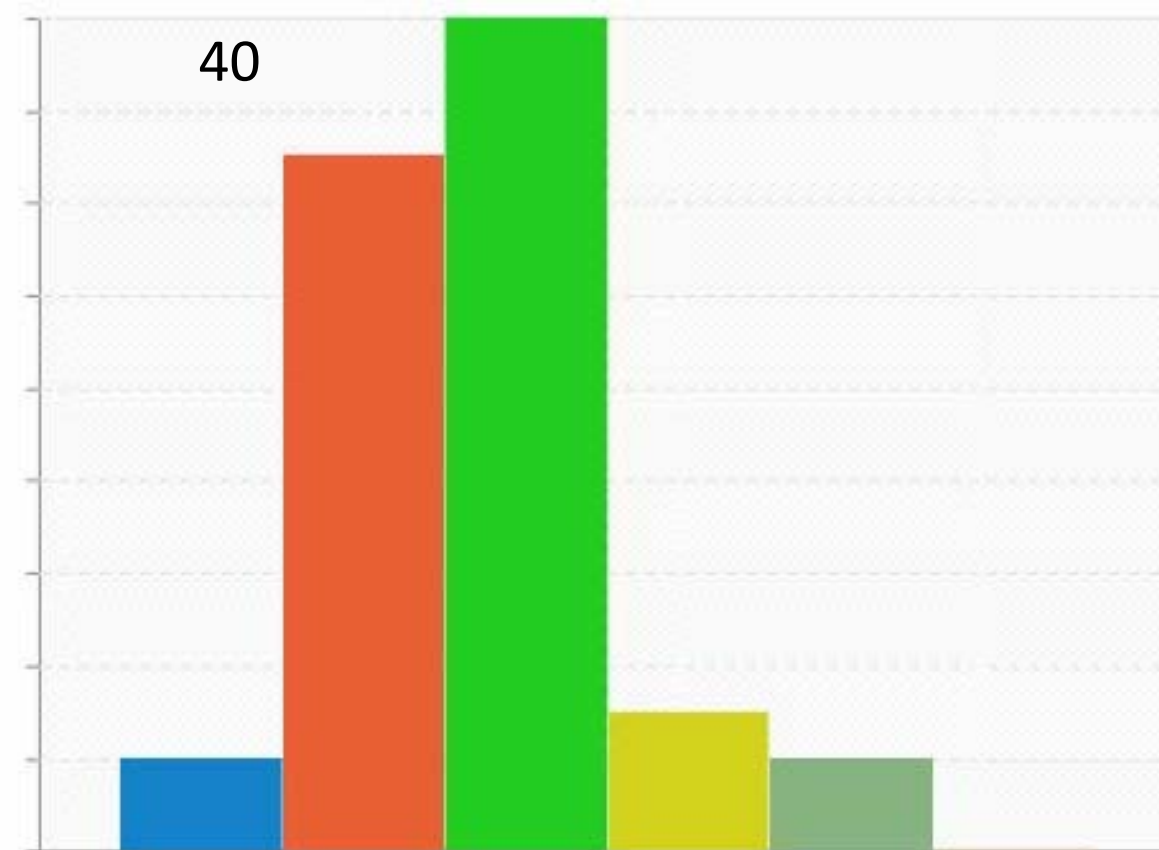


- Image pattern recognition (e.g. event reconstruction) (23)
- Classification (e.g. particle identification, physical processes) (22)
- Assignments (e.g. solving ambiguities with >2 identical particles) (6)
- Regression (calculation of variables, e.g. energy corrections) (23)
- Curiosity (currently no special research idea) (19)
- Applications are not obvious (1)
- Other (0)

Strong demands, expect exponentially increasing applications

GPU resources

I have access to GPU resources for network training at my research institute

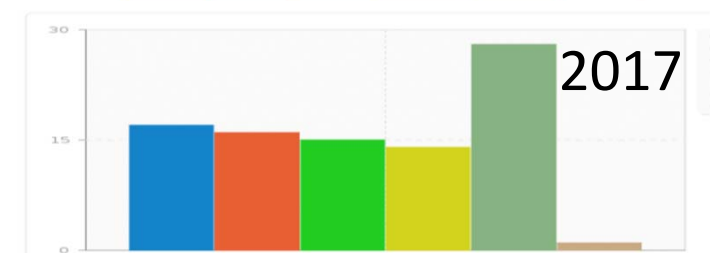
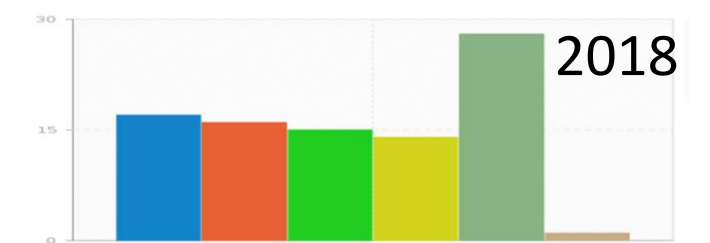
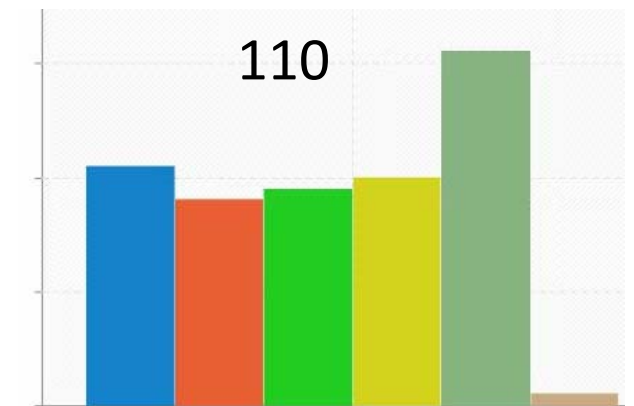
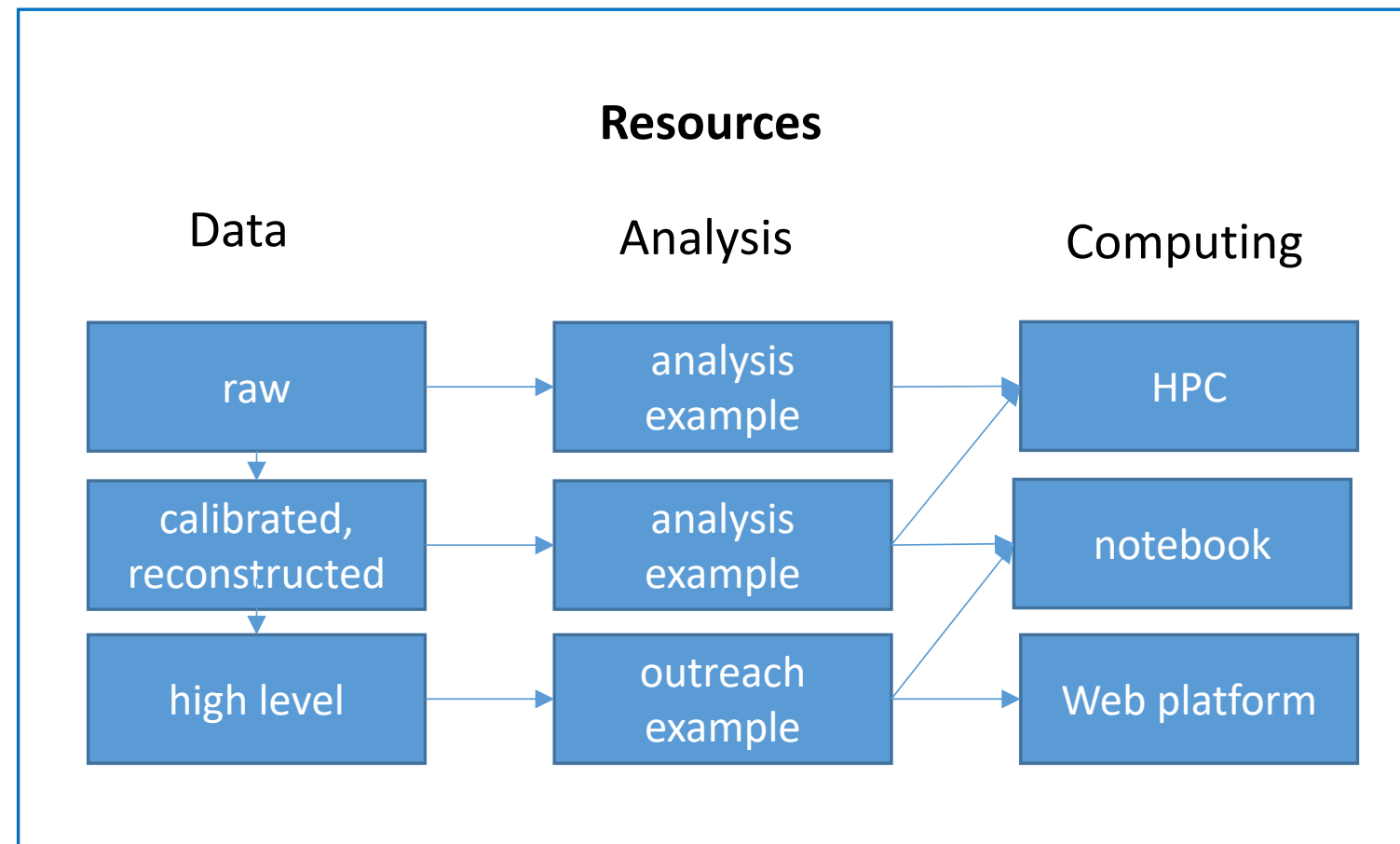


- None (3)
- Some (14)
- Sufficient (15)
- Don't know (4)
- No answer (0)
- Not displayed (0)

Experiment data, analyses, publications

For performing my research I need access to

- Raw data
- Calibrated data
- Reconstructed data
- High level data
- Simulated data
- Other



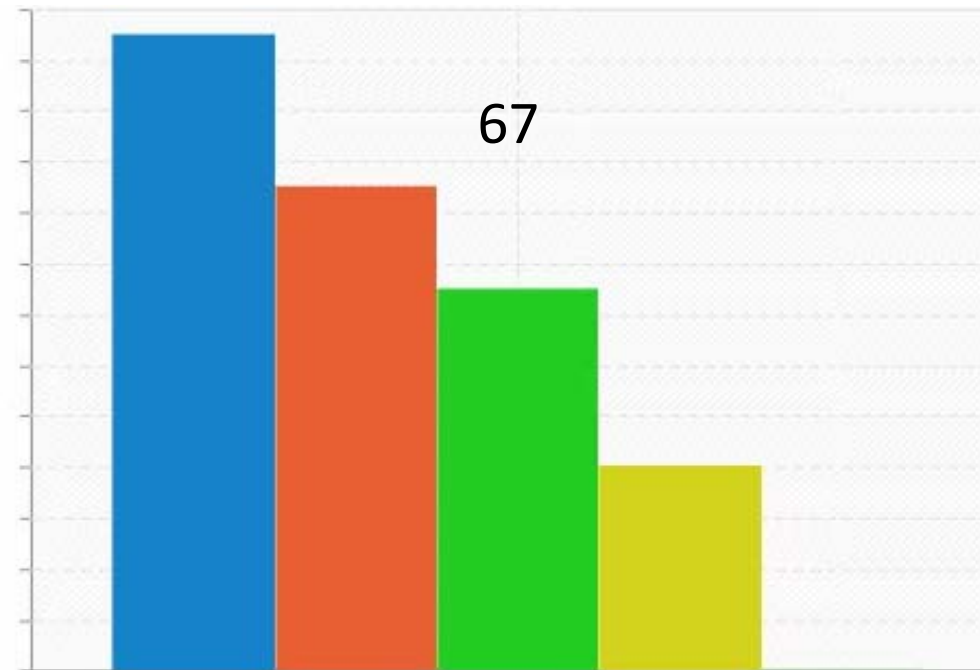
- Raw data (17)
- Calibrated data (16)
- Reconstructed data (15)
- High level data (direction, energy,...) (14)
- Simulated data (28)
- Other (1)

Resource intensive, aim to maximally exploit old & new data

Experiment data, analyses, publications

I get the most important analysis support from

- My research group/institute
- Colleagues of my collaboration
- Information on websites
- Previous analyses
- Other



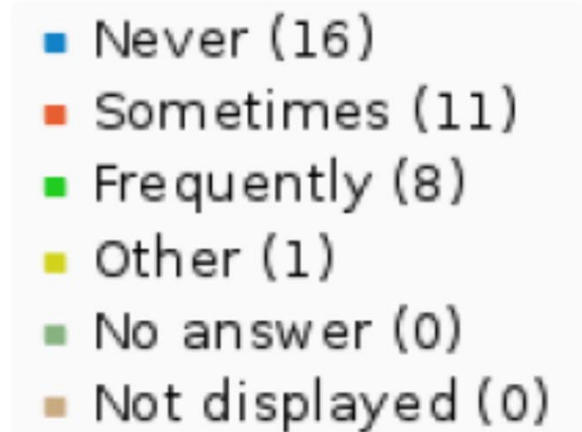
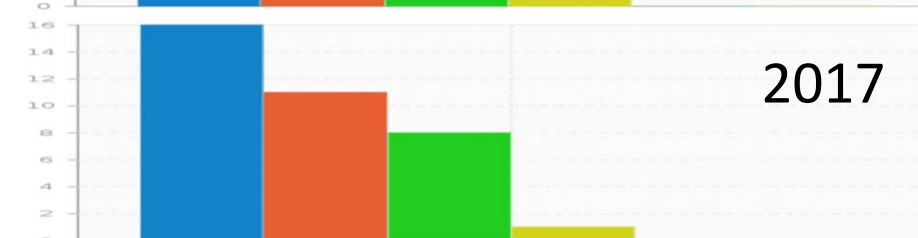
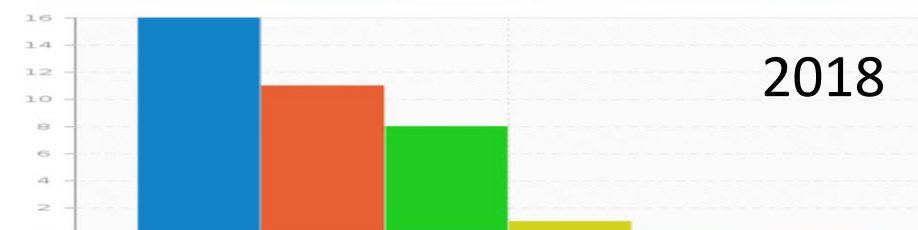
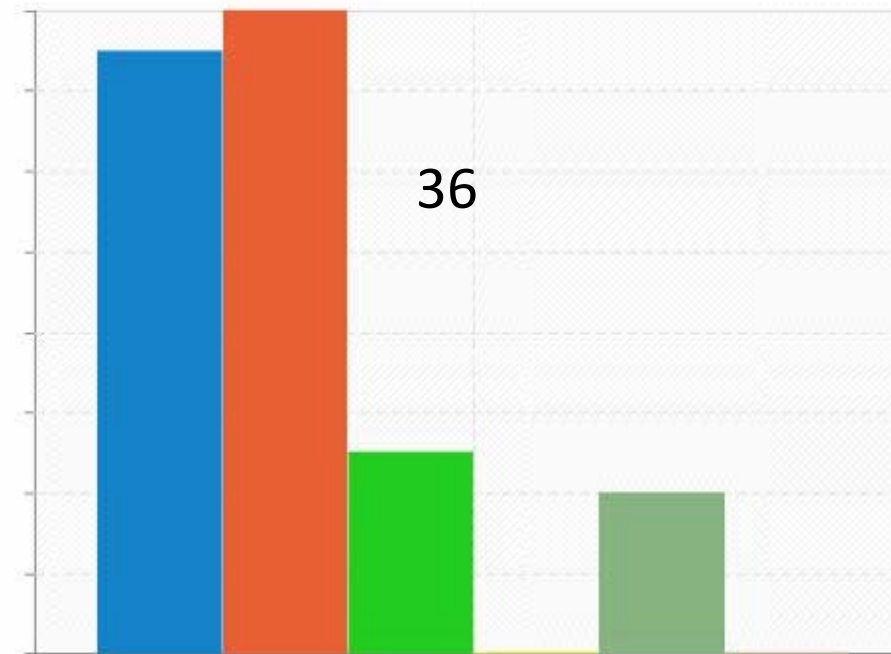
- My research group/institute (26)
- Colleagues of my collaboration (18)
- Information on websites (18)
- Previous analyses (10)
- Other (0)

Support is an obvious prerequisite for successful exploitation of data

Experiment data, analyses, publications

I have accessed open (*public*) data for my research purposes before

- Never
- Sometimes
- Frequently

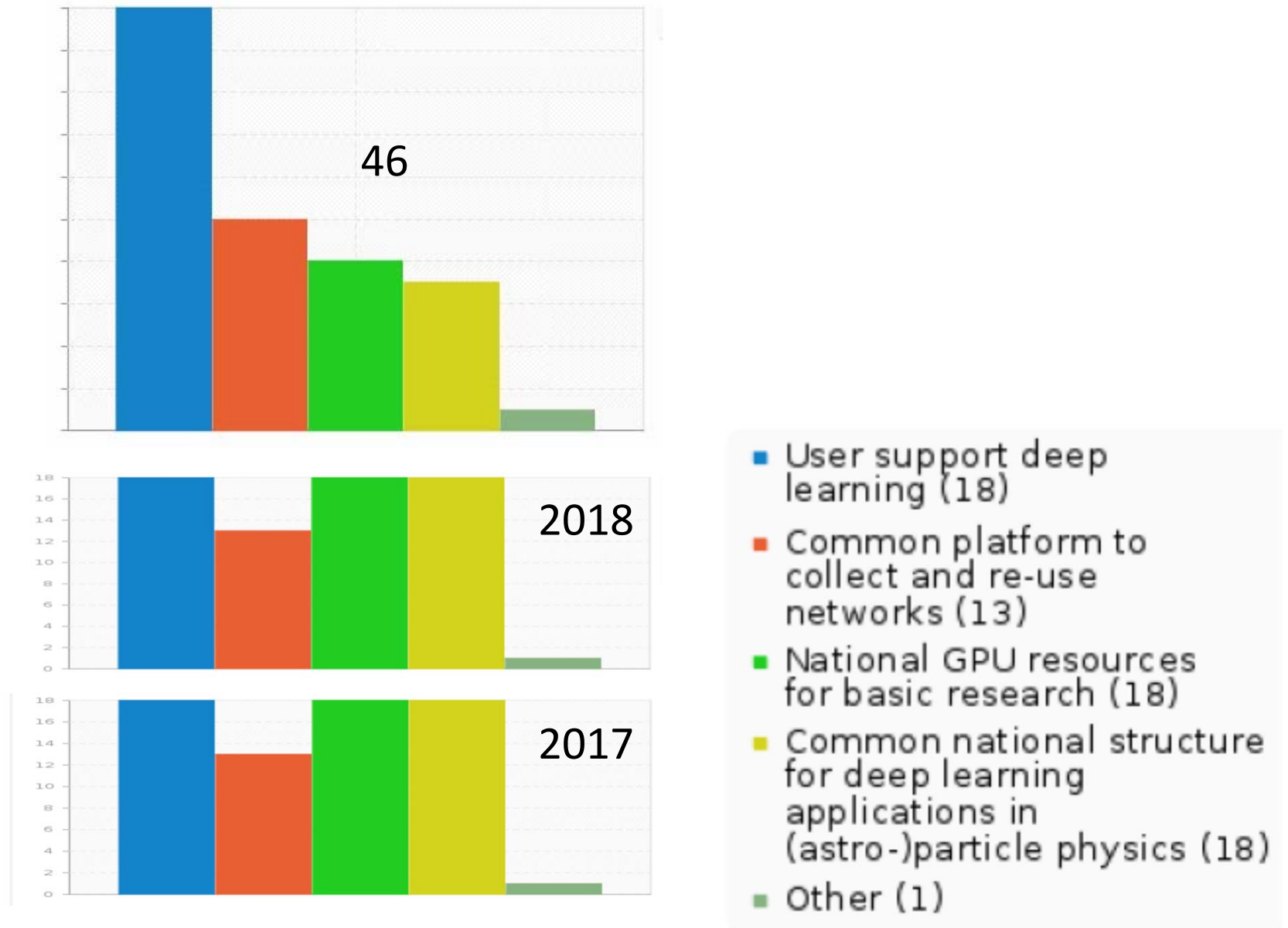


Requires scientific experience to judge what to extract from which data

Improve exploitation of new and old data

The following **measures** are important for achieving progress in my research

- User support deep learning
- Common platform to collect and re-use networks
- National GPU resources for basic research
- Common national structure for deep learning applications in (astro-)particle physics
- others

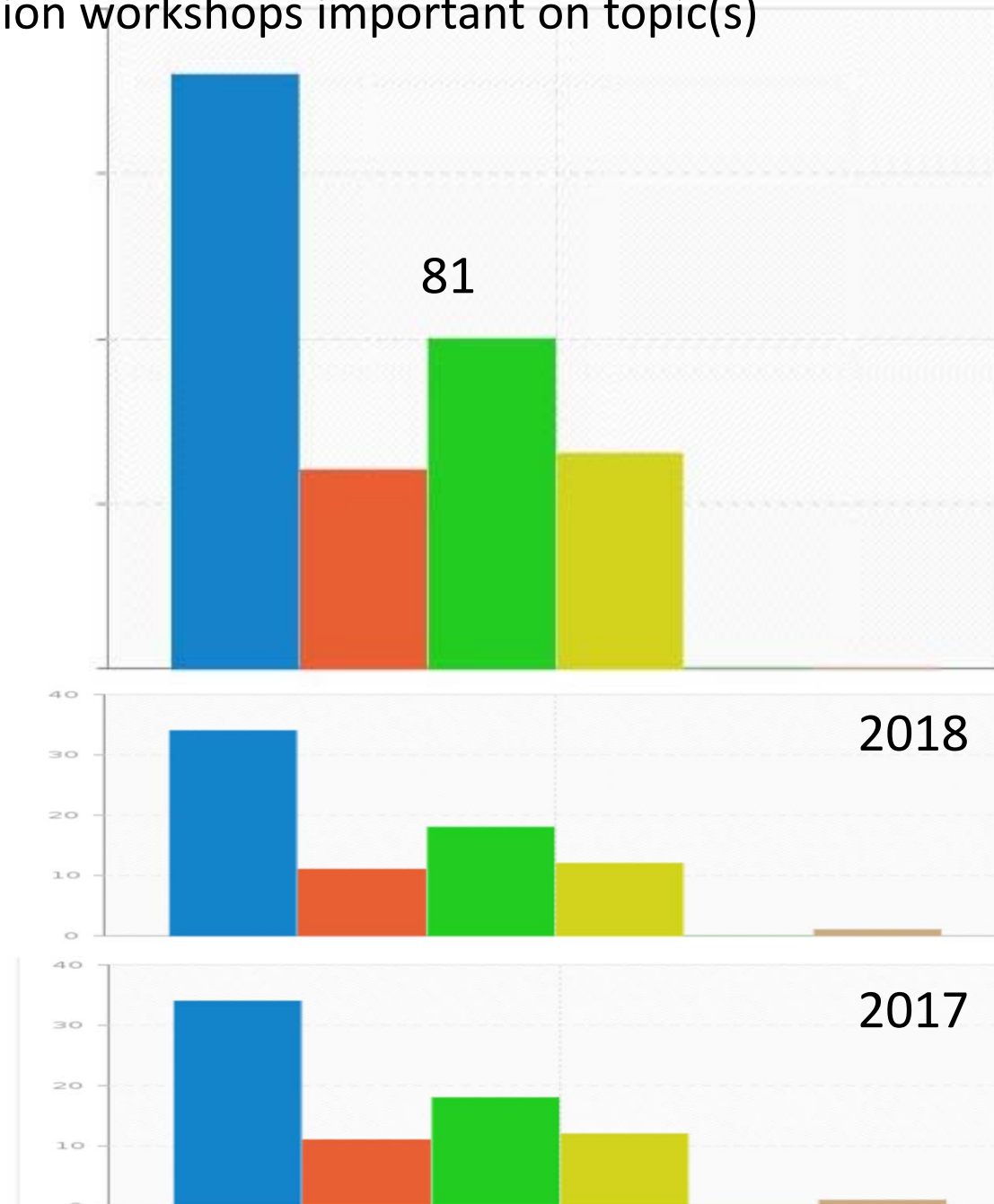


Need sustainable structures for accelerating research & development

Workshops perspectives

For the future, I consider continuation workshops important on topic(s)

- Deep learning
- Open data
- Open analysis software
- Analysis preservation
- Not important
- Other

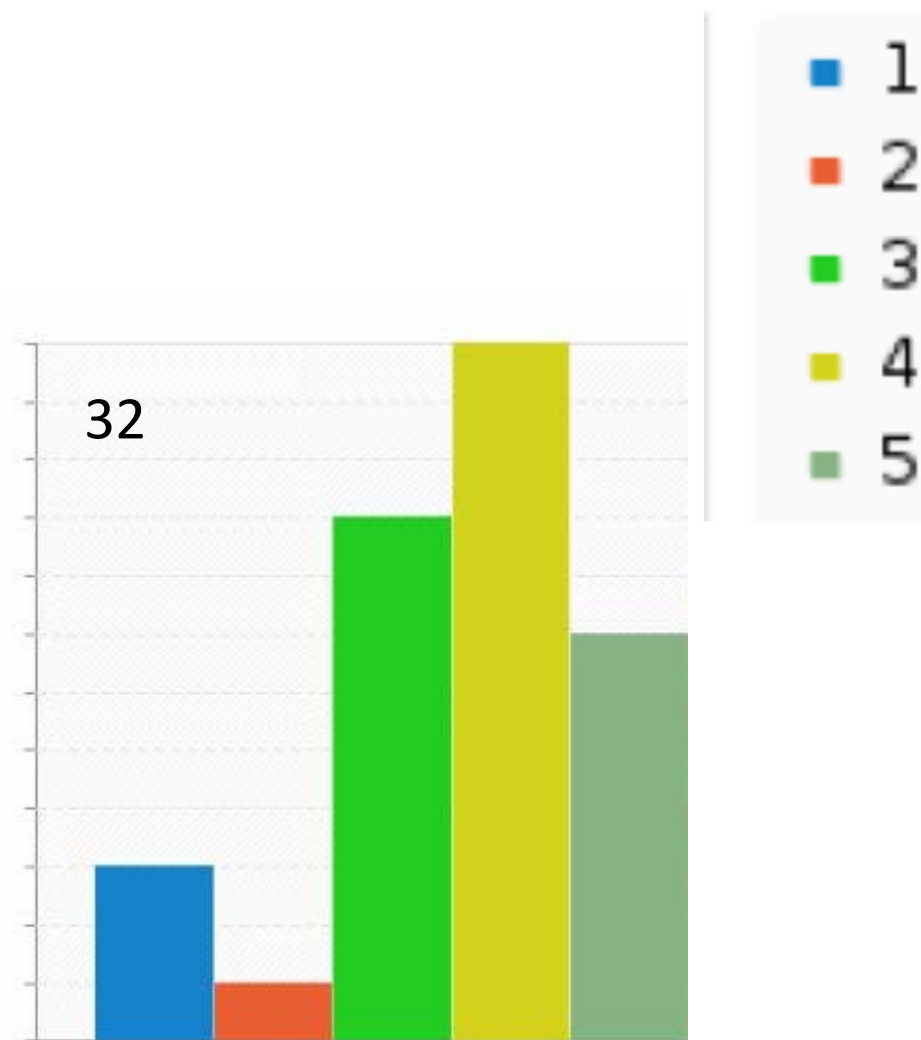


- Deep learning (34)
- Open data (11)
- Open analysis software (18)
- Analysis preservation (12)
- Not important (0)
- Other (1)

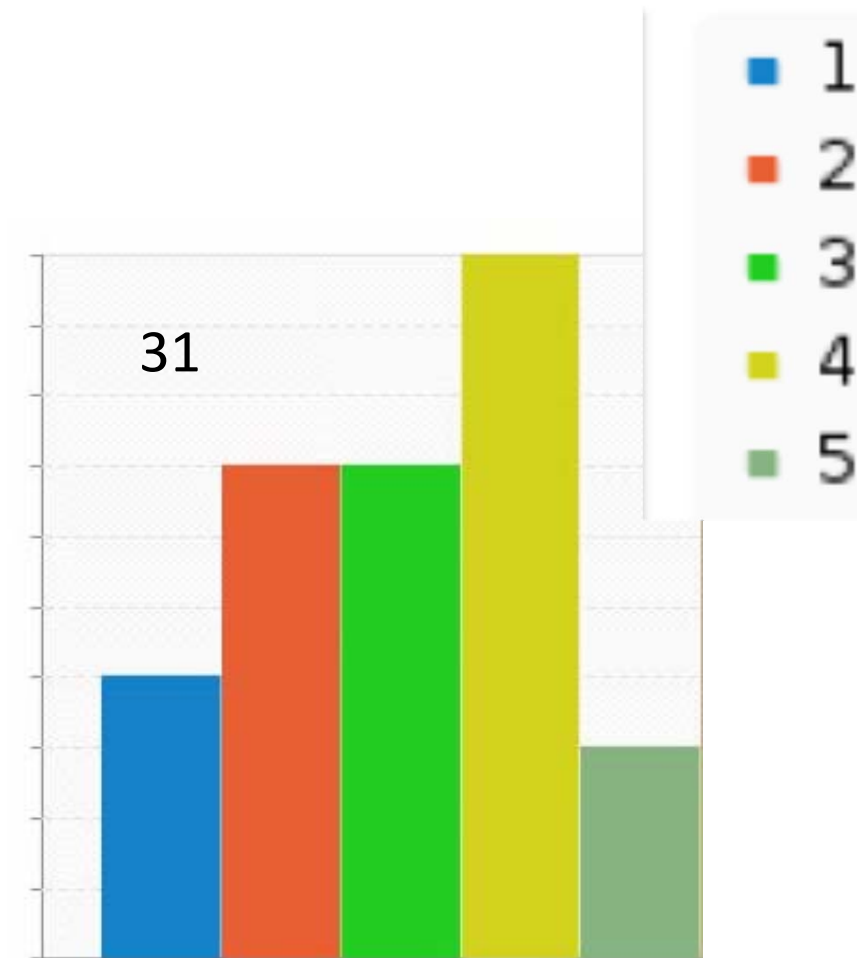
Strong wish to benefit from each other's progress

Workshop tutorial: *Information Field Theory*

The *IFT* introductory tutorial was helpful



The example applications deepened my understanding of *IFT*

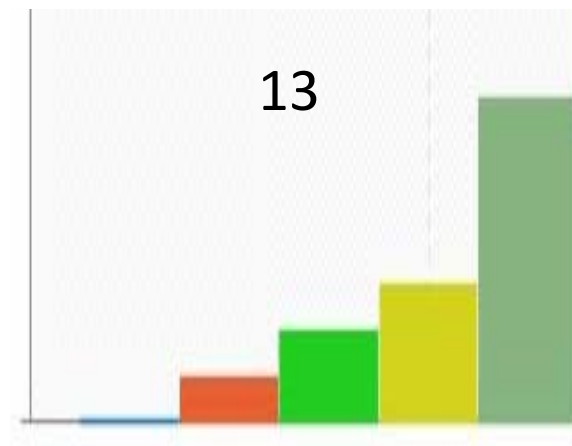


1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree

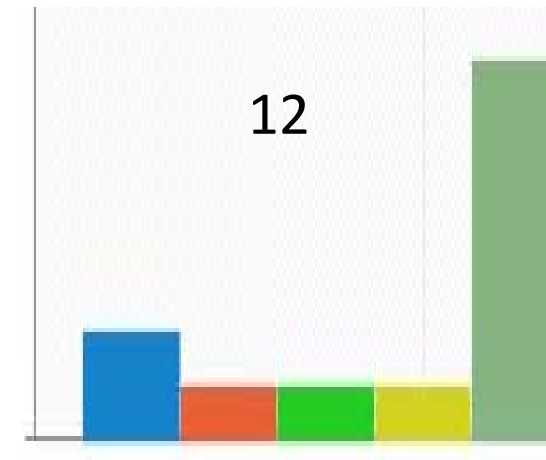
Workshop beginners tutorial

- Introduction to machine learning with deep neural networks

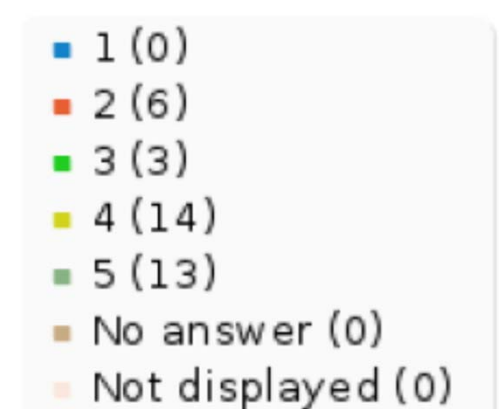
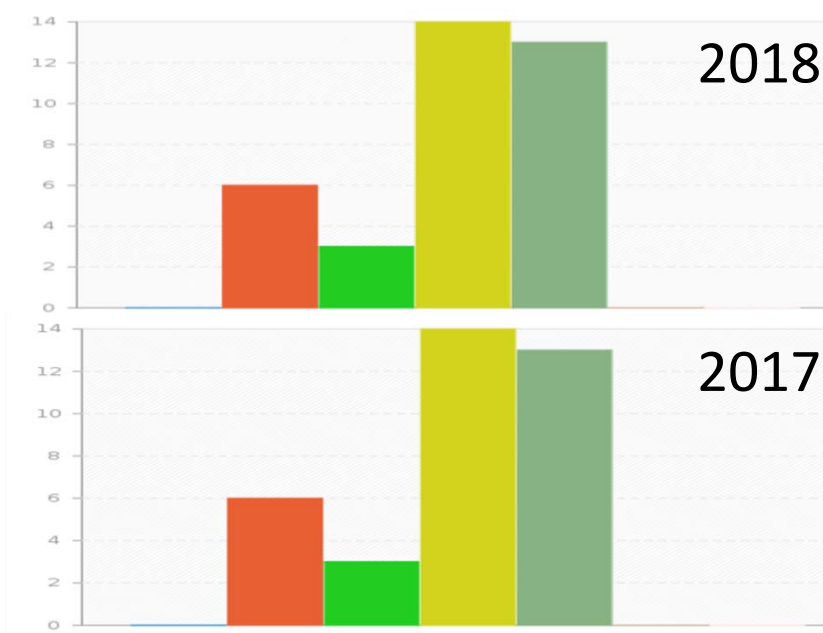
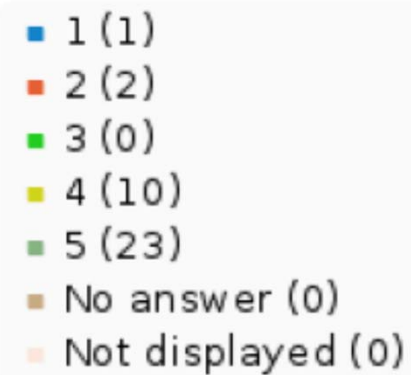
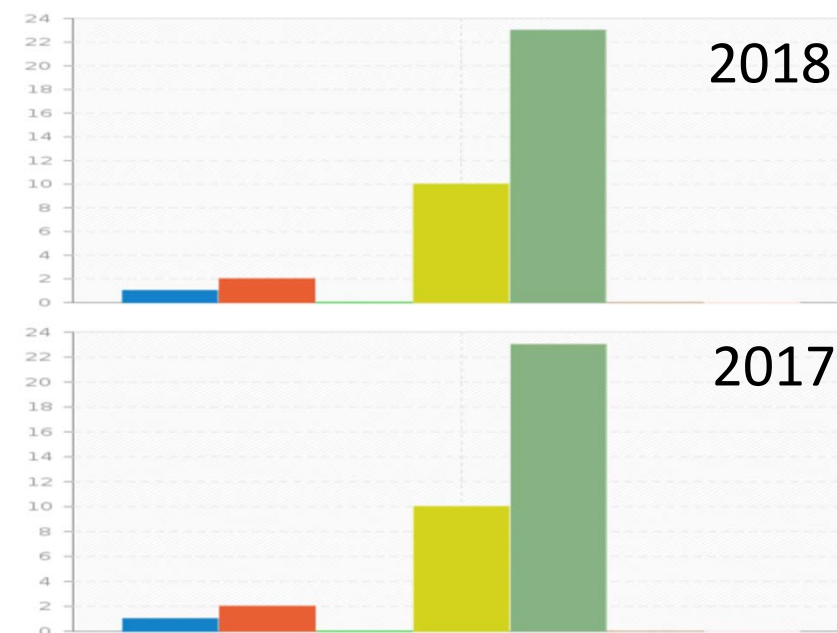
The deep learning introductory tutorial was helpful



The example applications deepened my understanding of deep networks



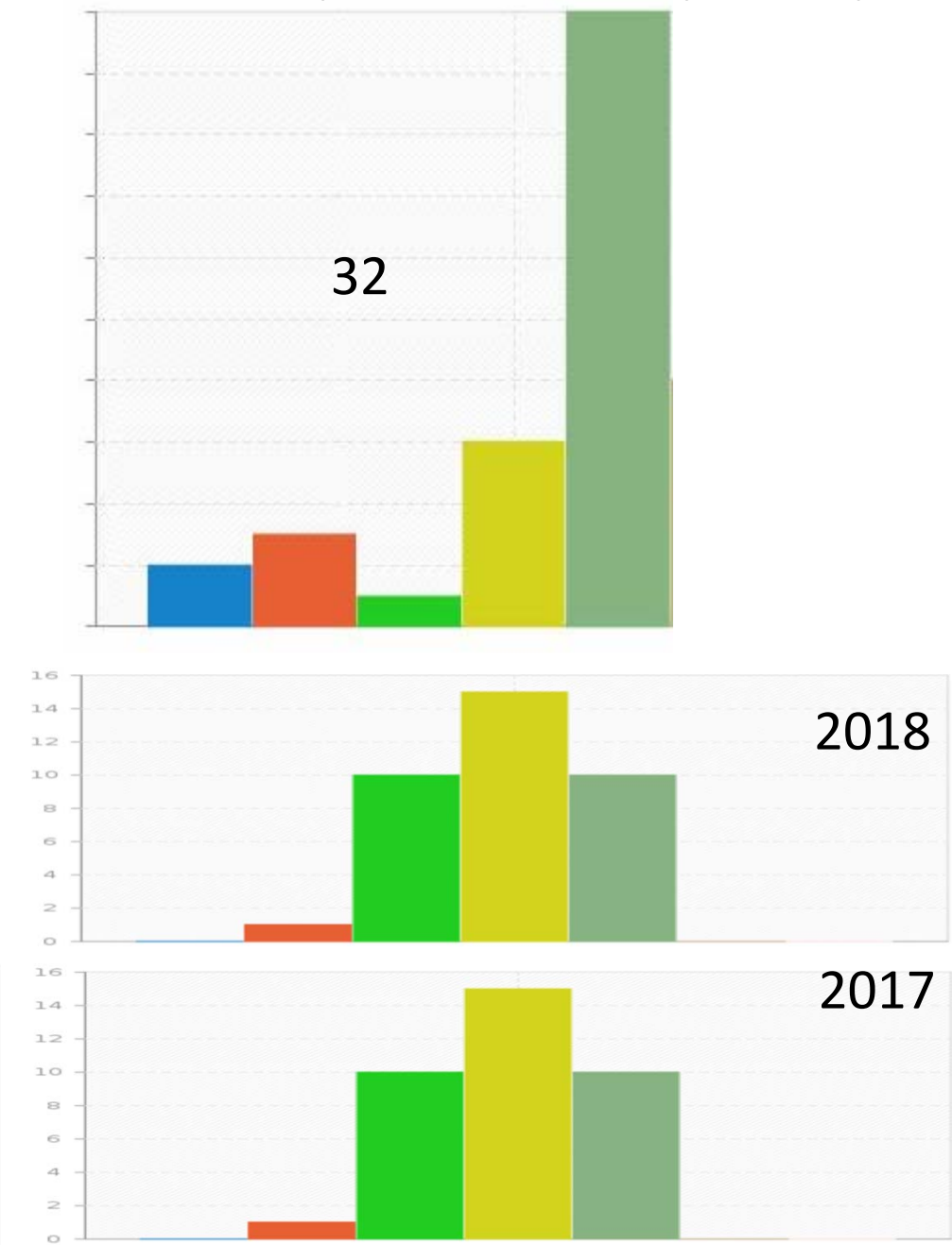
1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree



VISPA platform

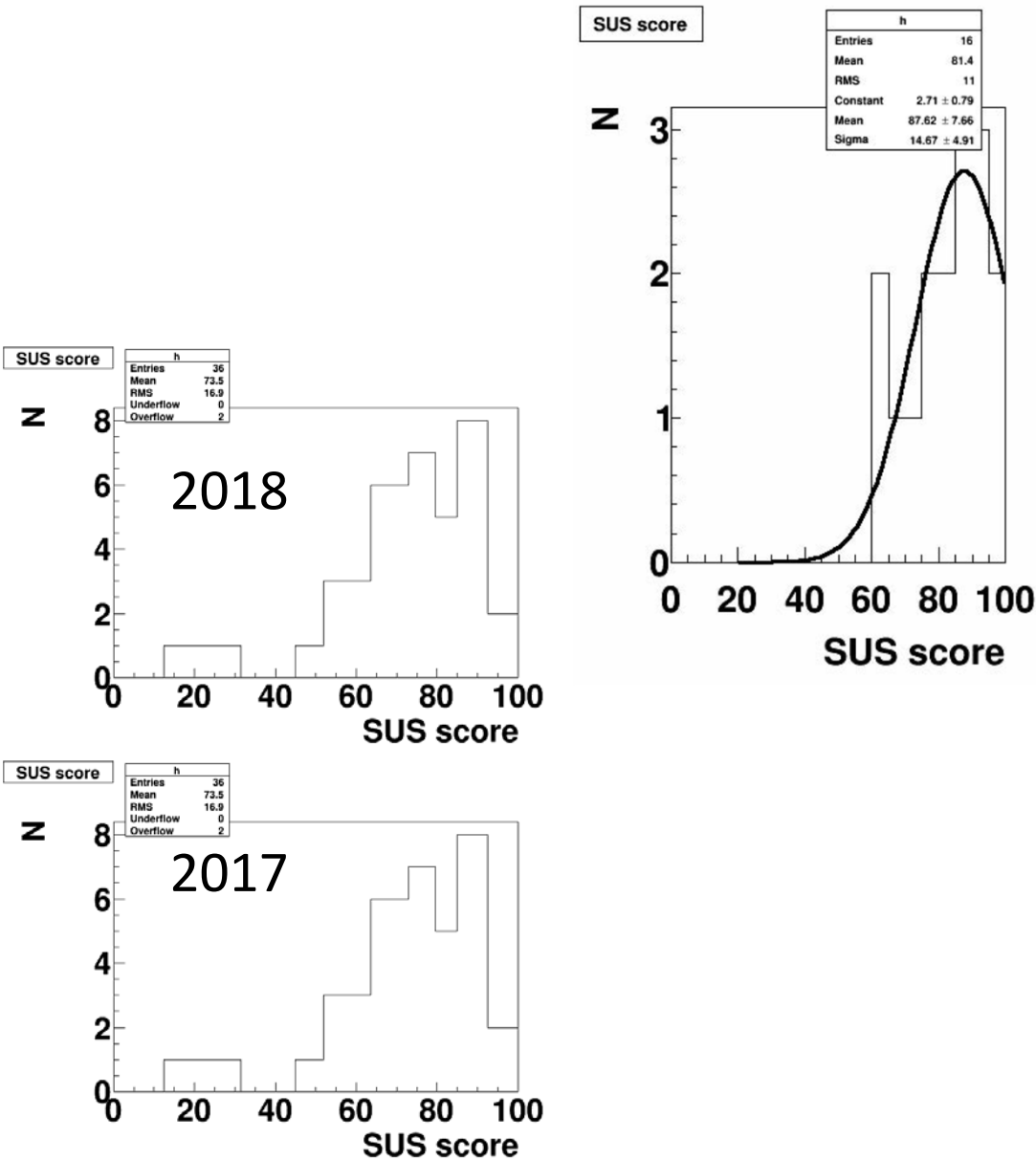
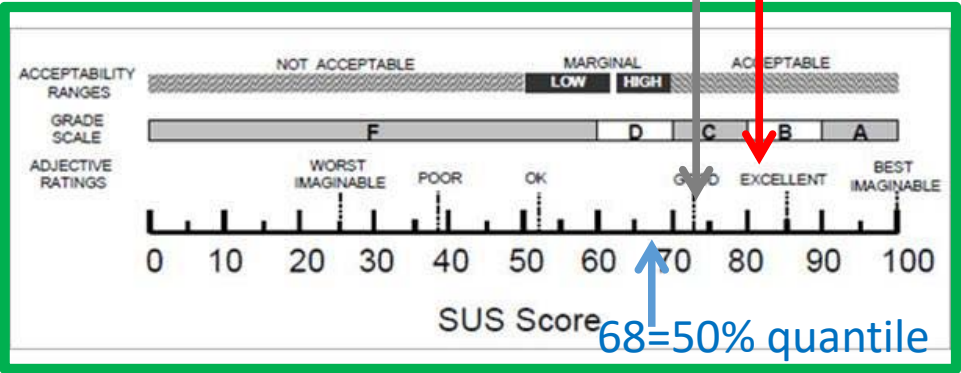
My overall score for the VISPA platform

1=insufficient, 2=sufficient, 3=ok, 4=good, 5=very good



- 1 (0)
- 2 (1)
- 3 (10)
- 4 (15)
- 5 (10)
- No answer (0)
- Not displayed (0)

Non-linear quality scale



Comments

More methodology talks and less result talks. Workshops are a learning platform. Showing accuracy scores of a single machine learning architecture doesn't teach me much and should be outsourced to classical conferences.

Some parts of the graphical user interface of vispa were slow, e.g. it crashed when browsing the directory of all users. Otherwise it was a great platform for tutorials.

Thanks a lot for this workshop. I personally think that the workshop was / is very successful and is motivating myself to discover more and more the topic of neural networks and deep learning. I could imagine to come next year as well and I think that it is of major importance to go on with these workshops.

Quality of conference dinner was really bad (*comment: please specify*)

From the discussion

- To emphasize the methodical part even more as a workshop.
- To take everyone even better, the first 90 seconds should be used to name the physical context (as everyone does), and then again 90 seconds to explain why they wanted to use exactly this method.
- In the contributions of the participants the methodical part itself can be emphasized even more by showing code snippets (e.g. from Tensorflow), which is the crucial trick with which something worked.
- It's not a conference where you present each other's great transparencis, but a workshop where you can also discuss the problem cases where something didn't work and what prevents the others from spending time with the same mistakes.
- Key note speakers will continue to be invited to bring information and developments from the other communities into our community, as we have done so far.
- Ethics should also find a place at this workshop, where else can you learn about it, if not here.
- Structure of the workshop: "Tutorial - Lectures - Looking beyond your own nose" is still ok, all components are wanted.
- Participant contributions can also be technical.
- The size of the workshop is appropriate because you can still talk to anyone. Clear focus in a very broad portfolio of Efforts around the world.
- Time between workshops, communication channels are needed to stay in touch, to be able to ask questions.

Thanks very much!

Workshop administration:

[Sabine Bucher, Melanie Roder](#)

Local RWTH workshop team:

[Jonas Glombitza, Yannik Rath, Marcus Wirtz](#)

and the VISPA team:

[Benjamin Fischer, Niclas Eich, Dennis Noll, Peter Fackeldey, and more](#)