# Welcome to the school «Machine Learning for High Energy Physics»

Andrey Ustyuzhanin

20th June 2016

# Organizers, a short intro

〉 **Yandex (est. 1994)**

— World-wide search engine, leading position (~60%) in Russia

— CERN openlab partner since 2013

〉 **Yandex School of Data Analysis (est. 2007)**

— member of CERN LHCb & SHiP collaborations since 2014

〉 **Yandex Data Factory (est. 2014)**

〉 **Higher School of Economics**

— Computer Science faculty (founded by Yandex)

— LAMBDA (**LA**boratory of **M**ethods for **B**ig **D**ata **A**nalysis)

# HEP Challenges

› Online event selection (10s TB/sec), data storage optimization

› Automatic event reconstruction

– Reconstruction of tracks from hits, or higher-level properties

– Semi-supervised algorithms

– Parallelized execution (GPU, Xeon Phi, etc)

› Anomaly detection & prediction, i.e. data quality certification

› Machine Learning for hypothesis testing (new physics?)

› Systematic error estimation for parametrized prediction models

› Fast event simulation

› High luminosity LHC «era»: 100x increase of data flow in 2025

Andrey Ustyuzhanin

# Machine Learning advances & tech

❯ Algorithm ensembling

❯ Deep learning

❯ Feature extraction automation

❯ Representation learning

❯ Transfer learning

❯ Clustering

❯ Outlier detection

❯ Collaborative filtering

❯ …

Andrey Ustyuzhanin

# Why MLHEP?

> Machine Learning: powerful approaches for complex tasks

> HEP: lots of challenges which could be solved with ML

> Increase people expertise in ML → advances in the field

> Help career path development

– TMVA citations: ~750

– scikit-learn citations: ~1760

# MLHEP focus

› Bring variety of ML methods & tools to HEP projects in systematic way

— Offline data analysis

— Trigger optimization

— Infrastructure optimization: data storage, data certification

› Foster communication between HEP & ML communities

› Find possibilities for new joint projects

› Improve science (both HEP & ML)

› Research Reproducibility

Andrey Ustyuzhanin

# Advanced Track

❯ For those who are familiar with basic concepts of ML

❯ Topics:

— Trigger system overview, machine learning in the LHCb topological trigger; speeding-up predictions for boosted decision trees and neural networks

— Tools and practices for reproducible research design and conduction

— Tracking approaches overview, methods and tools

❯ When: Monday - Wednesday, same timeslots for lectures/seminars as timeslots of the main track

❯ Where: Same building, Andromeda room

# Structure

❯ Two tracks (https://indico.cern.ch/event/497368/timetable/ )

❯ Invited Lectures

❯ Two challenges running this week

— Start today. Award ceremony on Sunday, 26th or June

❯ **Food & Cultural experience**

— Cultural day, Friday 24th of June

— Welcome dinner, coffee breaks, lunches, **closing dinner**

# Infrastructure

〉 **MLHEP cloud #1 (main)**

– 16 machine x 16 CPU cores x 96 GB RAM

– Docker containers for participants

– Ubuntu with all ML necessary libraries

– https://2016.mlhep.yandex.net/

– authentication by github

〉 **MLHEP cloud #2 (GPU)**

– provided by National Supercomputer Center of Finland
https://www.csc.fi/

– instructions will be available around Wed

Andrey Ustyuzhanin

9

# MLHEP Team

- Alexey Rogozhnikov (YSDA, HSE)

- Tatiana Likhomanenko (YSDA, HSE)

- Nikita Kazeev (YSDA, MIPT)

- Alexander Panin (YSDA)

- Mikhail Hushchyn (YSDA, MIPT)

- Irina Gerhart (HSE)

- Cate Doglioni (Lund University)

- Martin Ljunggren (Lund University)

- Andrey Ustyuzhanin (YSDA, YDF, HSE)

# MLHEP highlights

› **Experimental to certain degree**

— 2 tracks: separation is tentative

— Timetable may change! stay tuned
https://indico.cern.ch/event/497368/

› **School ML challenges**

› **Get help:**

— https://gitter.im/yandexdataschool/mlhep2016

— mlhep2016@yandex.ru

› **All feedback is welcome**

— twitter (#mlhep2016), email, ..., whatever!

Andrey Ustyuzhanin

# Instead of Conclusion

> Machine Learning: powerful approaches for complex tasks

> HEP: lots of challenges which could be solved with ML

> Increase people expertise in ML → advances in the field

> Interdisciplinary research - points of growth

> Challenge: make your research cooler with ML tools

Andrey Ustyuzhanin

Dr. Andrey Ustyuzhanin

head of CERN-Yandex
R&D group

twitter @AnaderiRu

anaderi@yandex-team.ru

# Thank you and welcome!