

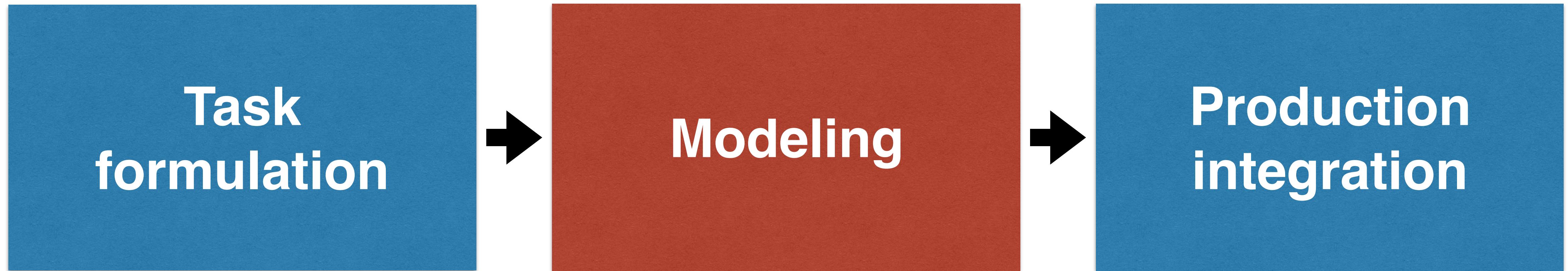
Competitive Data Science

@Tel Aviv

Dmitry Dremov

based on Peter Romov “Data Fest 3” slides

Data Scientist work scheme



Competitions platforms



DRIVENDATA

 **datascience.net**
 **CodaLab**

CORTANA INTELLIGENCE
Competitions




CIKM 25th ANNIVERSARY 2016
Indianapolis
October 24-28, 2016



Participant motivation



Participant motivation



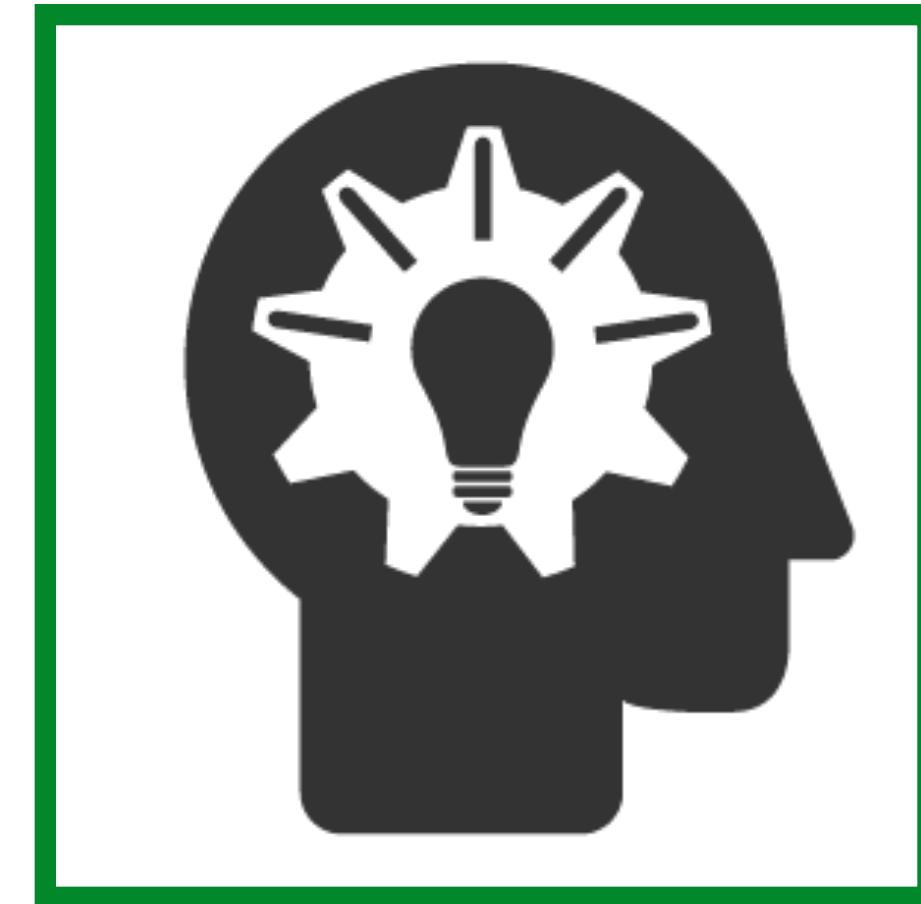
Participant motivation



Participant motivation



Participant motivation

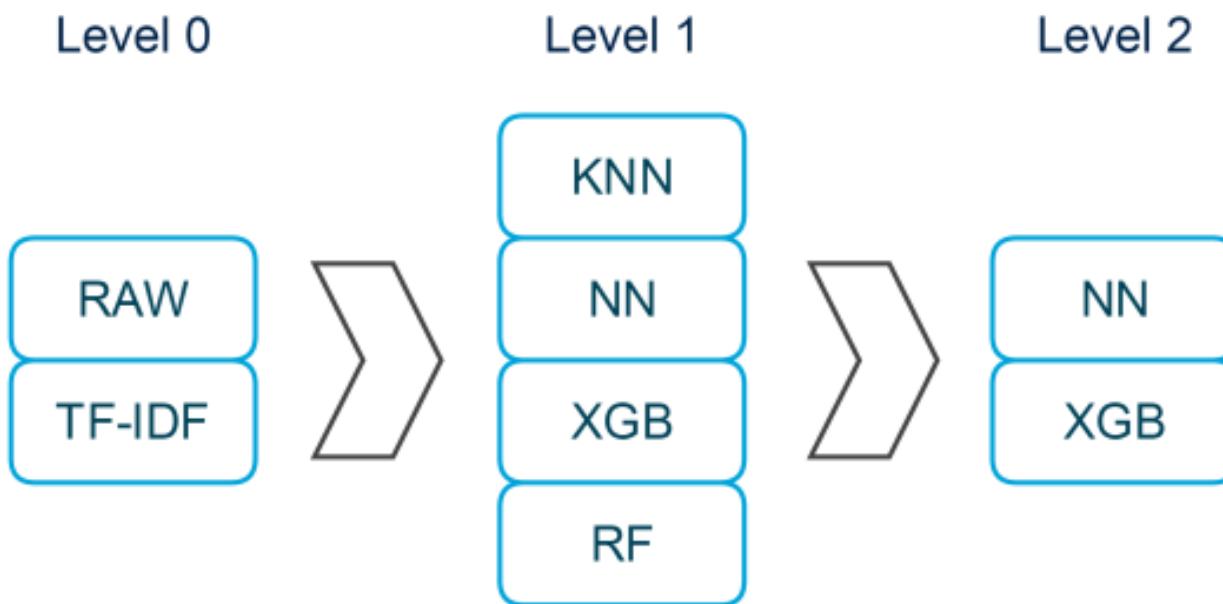


Competitions results

Problems attention



General
approaches



Tools

dmlc
XGBoost

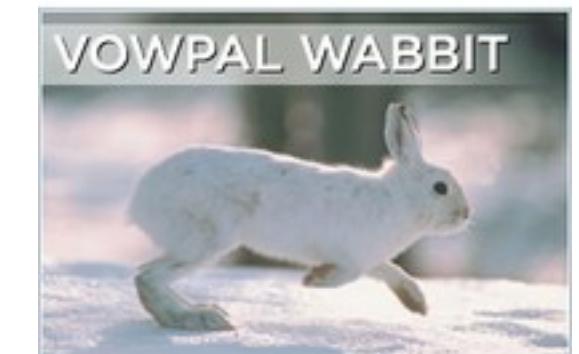


libFM

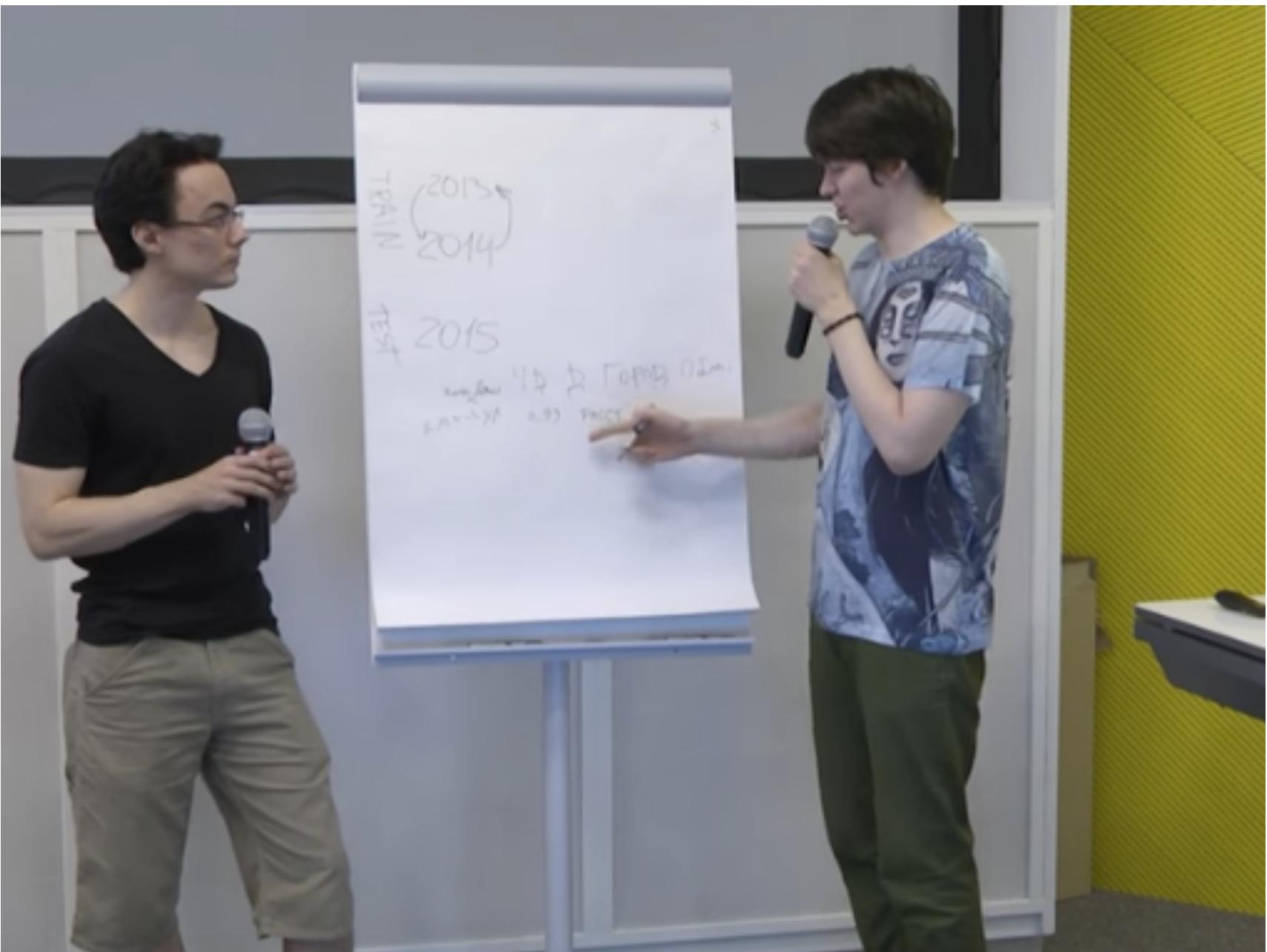


RGF

FTLR



Meetup format



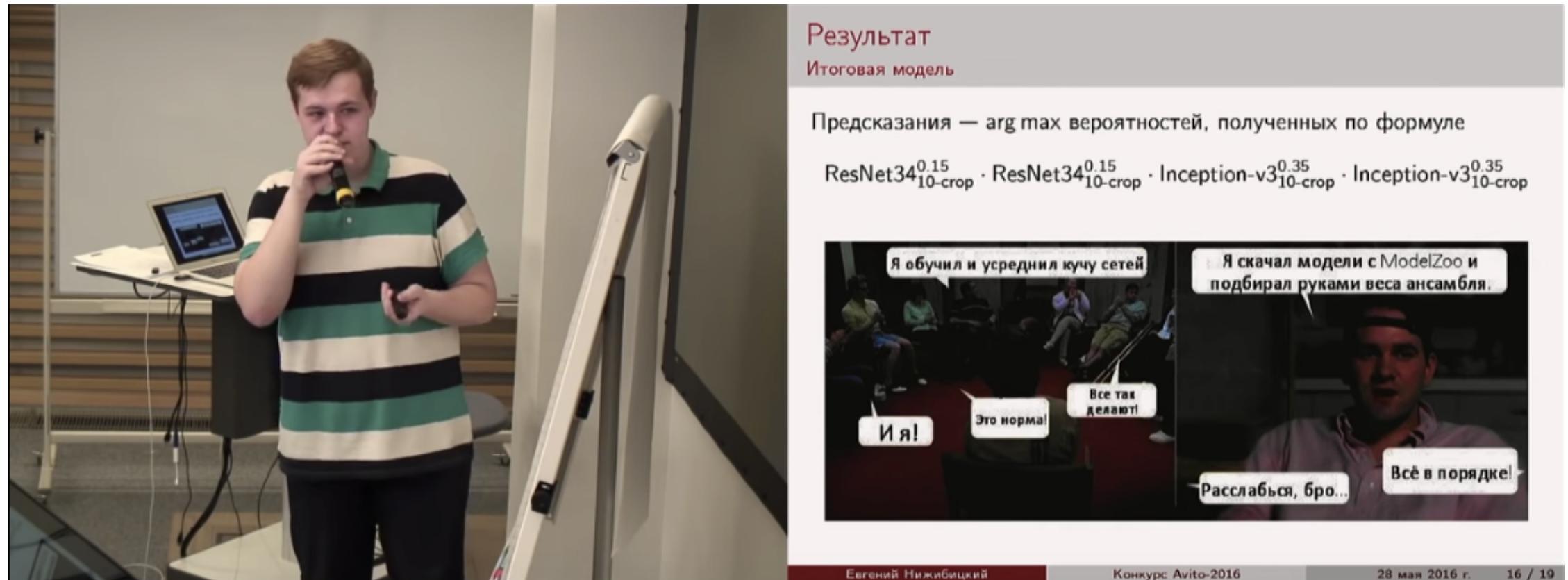
Solution sharing



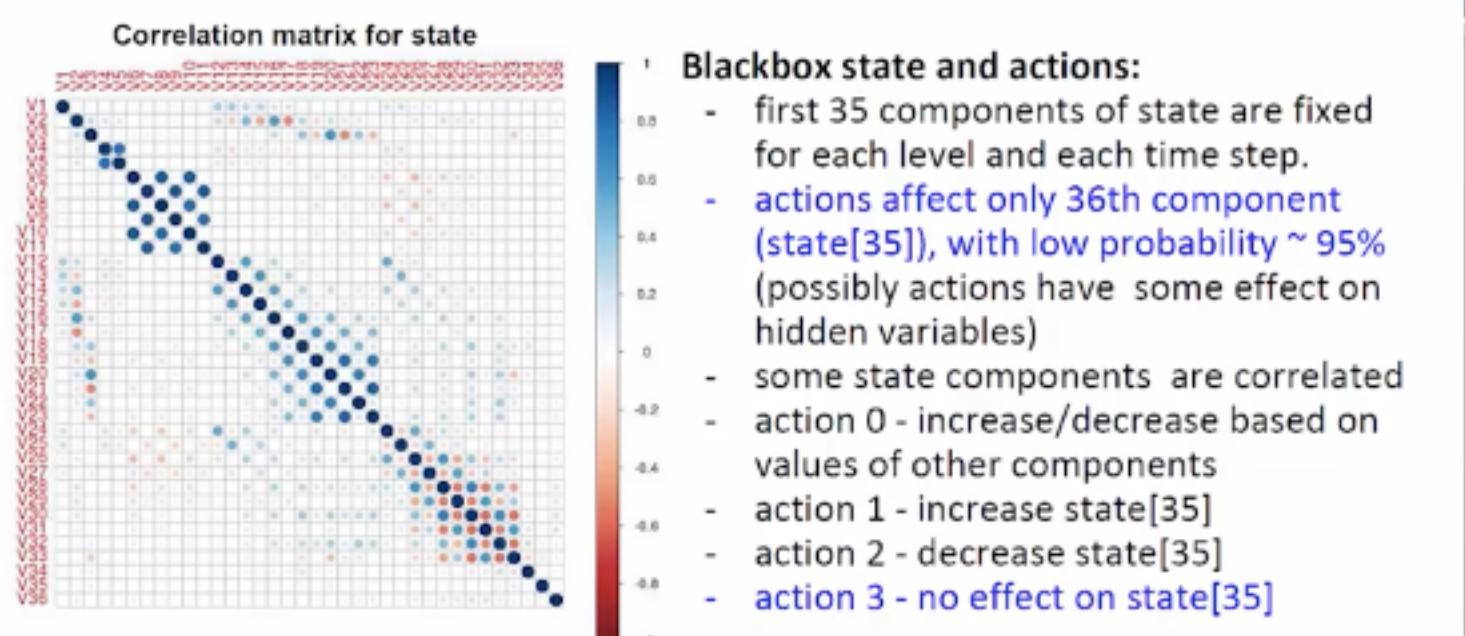
Hands-on work

Solution sharing

Winning and top solutions described by authors and community members



A. Black box problem. State and actions



Questions to speakers and authors

Discussion and networking during breaks

Hands-on work

- I want to start doing Kaggle!
- How to install XGBoost?
- What I can do, if server has not enough memory for dataset?
- How to validate model?
- I'm stuck. What else I can do?
- How to rent a server to grid search model parameters?
- Who wants to team up?
- ...

Competitions list on MLTrainings.ru

Ongoing competition from all platforms

Тренировки ML

Тренировки и разбор соревнований по анализу данных

[Facebook](#), [YouTube](#) [Сообщить о соревновании](#)

[Активны](#) [Завершены](#) [Все](#) [Поиск](#)

 **Cortana Intelligence Gallery: Women's Health Risk Assessment**
28 июня 2016 — 1 октября 2016
Осталось 3 недели, 1 день

The objective of this machine learning competition is to build machine learning models to assign a young woman subject (15-30 years old) in one of the 9 underdeveloped regions into a risk segment, and a subgroup within the segment. After the accurate assignments of the risk segment and subgroup in each region, a healthcare practitioner can deliver services to prevent the subject from the health risks, specifically sexual and reproductive health risks (like HIV infections). The types of services are personalized, based on the risk segment and subgroup assignments.

DIGINETICA
Codalab: CIKM Cup 2016 Track 2: Personalized E-Commerce Search Challenge
5 августа 2016 — 2 октября 2016
Осталось 3 недели, 2 дня

The Personalized E-commerce Search Challenge provides a unique opportunity for academia and industry researchers to test new ideas for personalized e-commerce search and consolidate the approaches already published and described in existing work. The successful participation in the challenge implies solid knowledge of learning to rank, log mining, and search personalization algorithms, to name just a few. For the model development, we release a new dataset provided by DIGINETICA and its partners containing anonymized search and browsing logs, product data, anonymized transactions, and a large data set of product images. The participants have to predict search relevance of products according to the personal shopping, search, and browsing preferences of the users. Both "query-less" and "query-full" sessions are possible. The evaluation is based on click and transaction data. The Challenge is a part

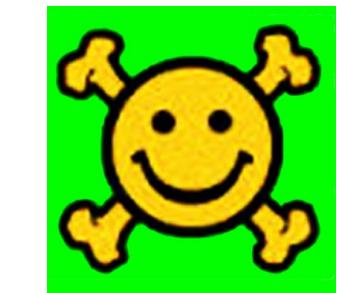
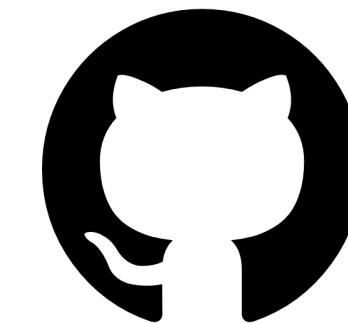
Recent meeting videos and slides (Russian)

 **BlackBox Challenge**
1 марта 2016 — 30 мая 2016 (завершено)

Вам предстоит научить агента принимать решения в условиях неопределенности в динамично меняющемся мире. Сможете ли вы победить в игре с неизвестными правилами, которые к тому же могут меняться со временем?

Разборы:

- Михаил Павлов, команда 5vision (2е место)   
- Павел Калинин (1е место) 



Thank you