

# Towards automation of data quality system for CERN CMS experiment

For the 22<sup>nd</sup> International Conference on Computing in High Energy and Nuclear Physics,  
CHEP 2016, October 10-14, hosted by SLAC and LBNL

Maxim Borisyak<sup>1,2</sup>, Jean-Roch Vlimant<sup>3</sup>, Andrey Ustyuzhanin<sup>1,2</sup>,  
Maurizio Pierini<sup>4</sup>, Maria Stenina<sup>5</sup>, Dmitry Smolyakov<sup>2</sup>

<sup>1</sup> National Research University Higher School of Economics; <sup>2</sup> Yandex School of Data Analysis;

<sup>3</sup> California Institute of Technology; <sup>4</sup> CERN; <sup>5</sup> Yandex

# Automated Anomaly Detection system

- › Daily operation of a large scale experimental setup is a challenging task both in terms of maintenance and monitoring.
- › Machine Learning-based Automated Anomaly Detection system can be trained online on expert decisions.
- › Trained model can assist data quality managers by filtering with obvious cases.
- › Experiments on CERN open portal data show that up to 60% of person power can be saved within reasonable increase in pollution and loss rates.