Towards automation of data quality system for CERN CMS experiment

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

Maxim Borisyak$^{1,2}$, Jean-Roch Vlimant$^3$, Andrey Ustyuzhanin$^{1,2}$, Maurizio Pierini$^4$, Maria Stenina$^5$, Dmitry Smolyakov$^2$

$^1$ National Research University Higher School of Economics; $^2$ Yandex School of Data Analysis; $^3$ California Institute of Technology; $^4$ CERN; $^5$ 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.