

Data Quality Monitoring with Machine Learning at CMS

Aytaj Aghabayli

Supervisors:

Maurizio Pierini

Jean-Roch Vlimant

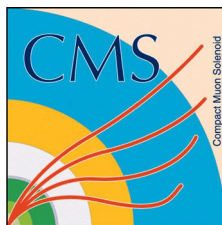
Who am I?

Aytaj Aghabayli

- Bs Computer Science

- Baku State University

- CERN CMS group



Yandex

Data Quality Monitoring?

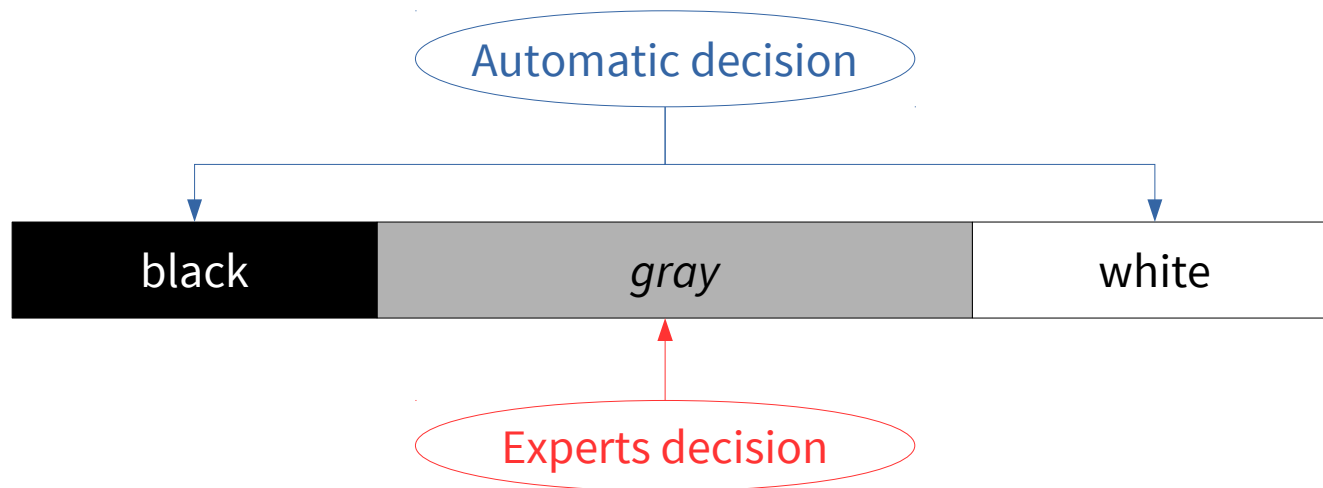
- Monitors and ensures data quality on each data
- Measuring data properties
- Anomaly detection
- Certification

DQM system at CMS

- Automation of DQM
- High Quality data for physics
- 'GOOD' and 'Bad' data
- Splitting data to 'black', 'white' and 'gray' parts

Data Quality Monitoring system at CMS

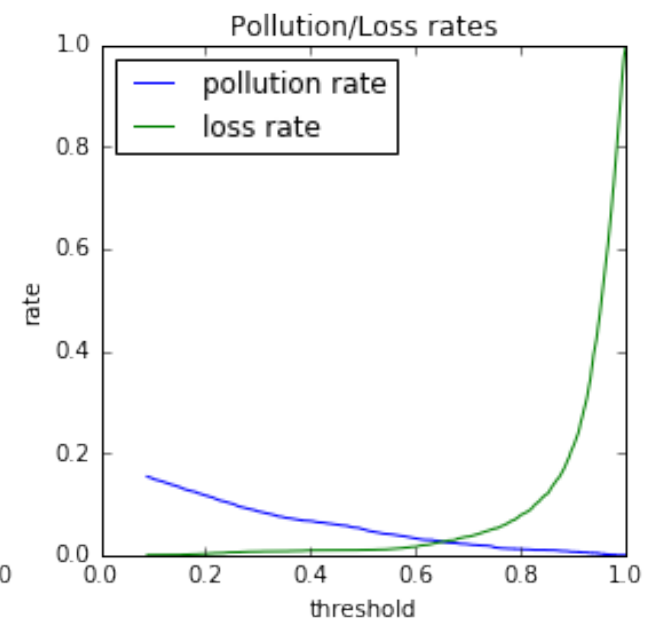
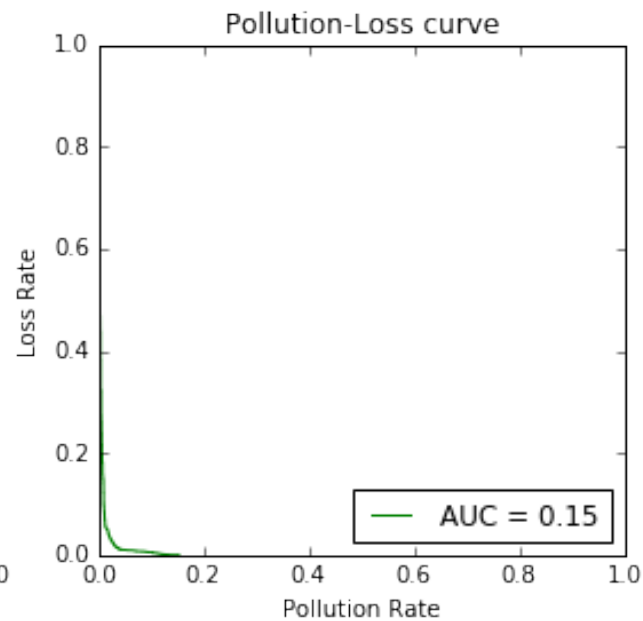
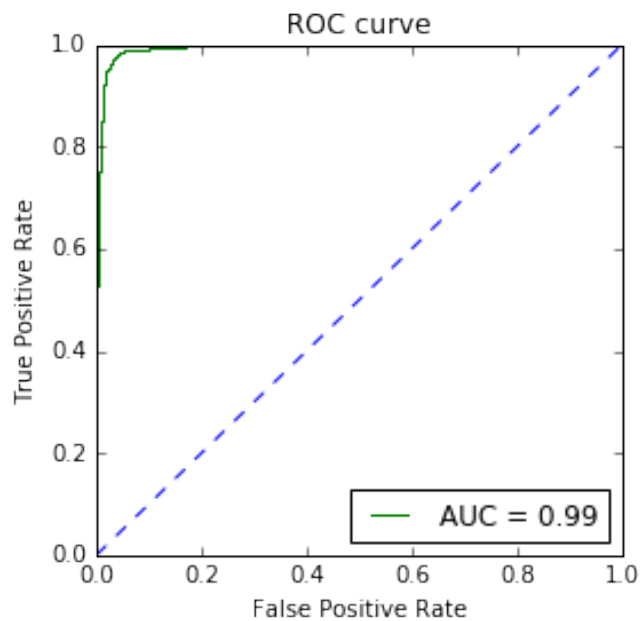
- “black zone” - LS is definitely bad
- “white zone” - LS is definitely good
- “gray zone” - classifier is in doubt, expert decision is needed



What have been done?

- 'Bad' JSONs
- Loading original data
- Extracted features ($p_T, \eta, \varphi, m, f_{x,y,z}$)
- Merged data
- Classified – Random Forest Classification
- Roc curves

Classification



Future work

- Applying the DQM system to 2016's year data
- New features
- Different classification methods
- Multi-class classification



Thank You!

:)

Aytaj Aghabayli

aytaj.aghabayli@gmail.com

<https://www.linkedin.com/in/aytaj-aghabayli-14632aa2>

