11th International Workshop on

DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS

Druskininkai, Lithuania, Hotel "Europa Royale"
http://www.mii.lt/DAMSS

November 28–30, 2019

Co-Chairmen:

Dr. Saulius Maskeliūnas (Lithuanian Computer Society)
Prof. Gintautas Dzemyda (Vilnius University, Lithuanian Academy of Sciences)

Programme Committee:

Prof. Juris Borzovs (Latvia)
Prof. Albertas Čaplinskas (Lithuania)
Prof. Robertas Damaševičius (Lithuania)
Prof. Janis Grundspenkis (Latvia)
Prof. Janusz Kacprzyk (Poland)
Prof. Ignacy Kaliszewski (Poland)
Prof. Yuriy Kharin (Belarus)
Prof. Tomas Krilavičius (Lithuania)
Prof. Julius Žilinskas (Lithuania)

Organizing Committee:

Dr. Jolita Bernatavičienė
Prof. Olga Kurasova
Dr. Viktor Medvedev
Laima Paliulionienė
Dr. Martynas Sabaliauskas

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt
Tel. +370 5 2109 315

Machine Learning for Data Quality Monitoring and Data Certification in the CMS Experiment at the LHC (CERN)

Mantas Stankevičius

Institute of Data Science and Digital Technologies
Vilnius University
mantas.stankevicius@mif.vu.lt

The Compact Muon Solenoid (CMS) is a general purpose detector working at the CERN Large Hadron Collider (LHC). Physics and detector status data are continuously accumulated with a rate close to 1 kHz, making extremely difficult to perform a real time data monitoring with fine time granularity. Teams of shifters have the duty to check recorded data in order to spot possible anomalies in the detector to avoid using the affected data in Physics analyses. This procedure is referred as Data Certification (DC) and has a primary importance to obtain good Physics results. CMS has a Data Quality Monitoring (DQM) group aimed to accomplish this task efficiently. Machine Learning (ML) aided automated tools can analyse large volumes of data in close to real time and help to detect data quality flaws and failures with lower latency. Efficient anomaly detection with alarm capability helps to save an expensive running time and to collect more valuable physics data. This work summarizes the current state of ML techniques applied to DQM and DC, as well as the on-going efforts trying to greatly reduce the granularity of this process to a few tens of seconds taking advantage of Artificial Intelligence.