Contribution ID: **345** Type: **Poster**

Prototype of a cloud native solution of Machine Learning as Service for HEP

Friday, 8 July 2022 20:10 (20 minutes)

Nowadays Machine Learning (ML) techniques are successfully used in many areas of High-Energy Physics (HEP), e.g. in detector simulation, object reconstruction, identification, Monte Carlo generation. ML will play a significant role also in the upcoming High-Luminosity LHC (HL-LHC) upgrade foreseen at CERN, when a huge amount of data will be produced by LHC and collected by the experiments, facing challenges at the exascale. To favor the usage of ML in HEP analyses, it would be useful to have a service allowing to perform the entire ML pipeline (in terms of reading the data, training a ML model, and serving predictions) directly using ROOT files of arbitrary size from local or remote distributed data sources.
The MLaaS4HEP framework is an R&D project inside CMS providing such kind of solution. It was successfully validated with a CMS physics use case which gave important feedback about the needs of analysts. For instance, we introduced the possibility for the user to provide pre-processing operations, such as defining new branches and applying cuts.
To provide a real service for the user and to integrate it into the INFN Cloud, we started working on MLaaS4HEP cloudification. This would allow to use cloud resources and to work in a distributed environment. In this work we provide updates on this topic, in particular we discuss our first working prototype of the service. It includes an OAuth2 proxy server as authentication layer, a MLaaS4HEP server, an XRootD proxy server for enabling access to remote ROOT data, and the TFaaS service in charge of the inference phase. With this architecture the user is able to submit a ML pipeline, after being authenticated, using local or remote ROOT files simply using HTTP calls.

In-person participation

Yes

Primary author: GIOMMI, Luca (Istituto Nazionale di Fisica Nucleare)

Co-authors: SPIGA, Daniele (Istituto Nazionale di Fisica Nucleare); KUZNETSOV, Valentin (Cornell University); BONACORSI, Daniele (Istituto Nazionale di Fisica Nucleare)

Presenter: GIOMMI, Luca (Istituto Nazionale di Fisica Nucleare)

Session Classification: Poster Session

Track Classification: Computing and Data handling