

Open Source Machine Learning Software Development at CERN and High Energy Physics: Lessons and experience exchange

Sergei V. Gleyzer
University of Florida/
CERN
`sergei@cern.ch`

September 30, 2018

Abstract

In my talk I would like to share the experience and lessons learned in creating an open source machine learning software at CERN with a significant user base in physics, expanding the open-source initiative to the wider physics community and our interactions with the open source community at large.

1 About the author

I am currently the project leader of the open-source Toolkit for Multivariate Analysis which is part of the ROOT Data Analysis and Software framework, the prevalent software used for particle physics data analysis (Large Hadron Collider and other experiments, 20k+ scientists). I simultaneously hold various positions (senior researcher in the University of Florida, Fermilab distinguished researcher and visiting fellow at CERN). I am also the founder and administrator of the High-Energy Physics Software Foundation Google Summer of Code program, Google Summer of Code Umbrella Organization, which has expanded significantly during my leadership (2016-) from 11 to 29 students working on open-source software development for high-energy physics.

Some relevant links:

CERN-HSF Google Summer of Code webpage:
<https://hepsoftwarefoundation.org/activities/gsoc.html>

Toolkit for Multivariate Analysis in ROOT: <https://root.cern.ch/tmva>

TMVA Github: <https://github.com/root-project/root/tree/master/tmva>

My personal website: <http://sergeigleyzer.com>

1.1 MLOSS Workshop

I would like to describe and exchange the experience of creating a fully-open source machine learning project and its expansion in the community. I would like to also interact with others with similar experience. In the past three years I have led both the Toolkit for Multivariate Analysis project and the CERN-High-Energy Physics Software Foundation Google Summer of Code program (umbrella organization for high-energy physics open source software development). In the past two years we have expanded the program three-fold with 40+ organizations and 90+mentors and 29 students in 2018, about half of which have worked on OS machine learning projects. Although a resounding success in our domain, there is still a lot to learn for us from other domains. We have recently started to collaborate significantly with other open-source ML projects (OpenML/OpenCV) and would like to continue contributing to the open-source ML community at large.