Contribution ID: **254**   Type: **talk**

Implementation of Jupyter Notebooks into The Reproducible Open Benchmarks for Data Analysis Platform (ROB)

Monday 12 July 2021 16:45 (15 minutes)

The Reproducible Open Benchmarks for Data Analysis Platform (ROB) is a platform that allows for the evaluation of different data analysis algorithms in a controlled competition-style format [1]. One example for such a comparison and evaluation of different algorithms is the "The Machine Learning Landscape of Top Taggers"paper, which compiled and compared multiple different top tagger neural networks [2]. Motivated by the significant amount of time required to organize and evaluate such benchmarks, ROB provides a platform that automates the collection, execution, and comparison of participant submissions in a benchmark. Although convenient, the ROB currently requires participants to package their submissions into docker containers, which can pose an additional burden due to the steep learning curve.

To increase ease of use, we implement support for the commonly used Jupyter Notebooks [3] in ROB. Jupyter Notebooks are a popular tool that many physicists are already familiar with. Using Jupyter notebooks, physicists are able to combine live code, comments, and documentation inside one document. By utilizing the PaperMill package [4], we allow ROB users to submit their implementations directly as Jupyter Notebooks in order to evaluate different data analysis algorithms without the need to package the code into Docker containers. To demonstrate functionality and spur usage of the ROB, we provide demos using bottom and top tagging neural networks that display the application of the ROB within particle physics as a way of providing a competition style platform for algorithm evaluation [5].

References:
[1] "Reproducible and Reusable Data Analysis Workflow Server", https://github.com/scailfin/flowserv-core
[2] Kasieczka, Gregor, Plehn, Tilman, Butter, Anja, Cranmer, Kyle, Debnath, Dipsikha, Dillon, Barry M, . . . Varma, Sreedevi. (2019). The Machine Learning landscape of top taggers. SciPost Physics, 7(1), 014.
[3] "Jupyter Notebooks", https://jupyter.org/
[4] "Papermill", https://papermill.readthedocs.io/en/latest/
[5] "Particle Physics", https://github.com/anrunw/ROB

Are you are a member of the APS Division of Particles and Fields?

No

Authors: WANG, Aaron; RAWAT, Ajay (University of Washington); MUELLER, Heiko; HSU, Shih-Chieh (University of Washington Seattle (US))

Presenters: WANG, Aaron; MUELLER, Heiko

Session Classification: Computation, Machine Learning, and AI

Track Classification: Computation, Machine Learning, and AI