

# ML software usage in LHCb

Paul Seyfert  
collecting input from many LHCb colleagues

INFN Milano Bicocca

29th September 2016



I asked around in LHCb what people think of a ML software installation on CVMFS.

- power users and developers, developing their tools and installing their software in their home directories (and happy with that)
- “normal” users, interested in stable versions which are easy to use. Want to use some existing, tested functionality in an analysis (target group for cvmfs installation)

## developers

- e.g. `root_pandas`
  - will keep using private installations (needed to develop new code)
  - interest in picking the right/latest/unreleased versions of various packages
  - ⚡ Sometimes end up putting private installations of root on lxplus (through `root-conda-recipes`) to avoid compatibility problems
- ⇒ up to date python with virtualenv and a compatible root available from cvmfs is about what power users use.

## users

- interested in stable versions
  - don't want to fight with compatibility problems
  - prefer preinstalled software over building software themselves to get started
- ⇒ all wishes on the next slide come from users

## hep\_ml

- uBoost (1305.7248, used in [Phys.Rev.Lett. 115 \(2015\), 161802](#) )
- UGradientBoosting (1410.4140, used in [LHCb-CONF-2016-006](#))
- BDT reweighting

## XGBoost

- already used in the development for B flavour tagging in LHCb run 2 (with initial headache of how to get *one* stable xgboost version accessible to all users)

## Meerkat

See [Anton's talk in the last meeting](#)

## REP

"quite a useful extension to TMVA and sklearn"  
See [Andrey's presentation at the February IML](#)

## Spearmint

"Spearmint is a software package to perform Bayesian optimization."

## other python tools

[root\\_numpy](#) and [root\\_pandas](#) (both not strictly machine learning)

## lasagne

new developments in calorimeter reconstruction