

UNSUPERVISED LEARNING

IDEAS FOR SAMPLE PRODUCTION STRATEGY

Judita Mamužić, Roberto Ruiz
IFIC / CSIC - University of Valencia
29 May 2019

CERN OPEN DATA

Filter by experiment

<input type="checkbox"/> ALICE	15
<input checked="" type="checkbox"/> ATLAS	101
<input checked="" type="checkbox"/> CMS	878
<input type="checkbox"/> LHCb	3
<input type="checkbox"/> OPERA	831

Filter by year

<input type="checkbox"/> 2010	32
<input type="checkbox"/> 2011	449
<input type="checkbox"/> 2011-2012	1
<input type="checkbox"/> 2012	484

Filter by file type

<input type="checkbox"/> aod	85
<input type="checkbox"/> aodsim	681
<input type="checkbox"/> csv	13
<input type="checkbox"/> gz	2
<input type="checkbox"/> ig	94
<input type="checkbox"/> json	10
<input type="checkbox"/> raw	1
<input type="checkbox"/> root	825
<input type="checkbox"/> txt	1
<input type="checkbox"/> xls	1

Filter by collision energy

<input type="checkbox"/> 7TeV	448
<input type="checkbox"/> 8TeV	447

Filter by topic category

<input type="checkbox"/> B-physics	16
<input type="checkbox"/> BSM Higgs	65
<input type="checkbox"/> SM Exclusive	50
<input type="checkbox"/> SM Higgs	183
<input type="checkbox"/> SM Inclusive	48
<input type="checkbox"/> SM Systematic Variations	19

Filter by keywords

<input type="checkbox"/> masterclass	55
--------------------------------------	----

- CERN Open Data <http://opendata.cern.ch/search?page=1&size=20&type=Dataset>
 - CMS has 14M data ($\sim 10\text{fb}^{-1}$) events from data at 8 TeV (different triggers).
 - ATLAS has 1.4M data ($\sim 1\text{fb}^{-1}$) and $\sim 2\text{M}$ MC events (some background samples) <http://opendata.atlas.cern/extendedanalysis/datasets.php>.
 - ATLAS is preparing 10fb^{-1} of 13 TeV data+MC for outreach purposes (in approval procedure):
 - **Background MC** samples tt, siglet, Z+jets, W+jets, diboson; SM Higgs.
 - **Signal MC** samples Z', mono-Z, SUSY gg->ttttDM, tt->ttDM, N2C1->WZMET, C1C1->llvvMET, slepton prod, Graviton.
 - **Data** with 1L,2L,1had1leptau,2hadhadtau0L,1hadtau0L,1LFatJet,2photons preselection
 - Information on e,mu,tau,photon, small/large R-jets, MET, b-tag, trigger, truth, weight, systematics.
 - There is some restriction for publication of results, education purpose.

HL-LHC / FCC MC PRODUCTION STRATEGY

- MC production strategy for FCC and HL-LHC can be re-used:
 - <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HLHELHCWorkshop>
 - https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HLHEWG_MC
 - <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/FutureHadroncollider>
 - <https://github.com/FCC-hh-framework/EventProducer>
 - /eos/experiment/fcc/hh/generation/lhe
- <http://fcc-physics-events.web.cern.ch/fcc-physics-events/LHEevents.php>
 - **QCD:** di-jet + 012 jet, slicing in HT (Pt of leading jet), tri-jet, four-jet
 - (**Diboson:** VV+012 jets, slicing in HT)