

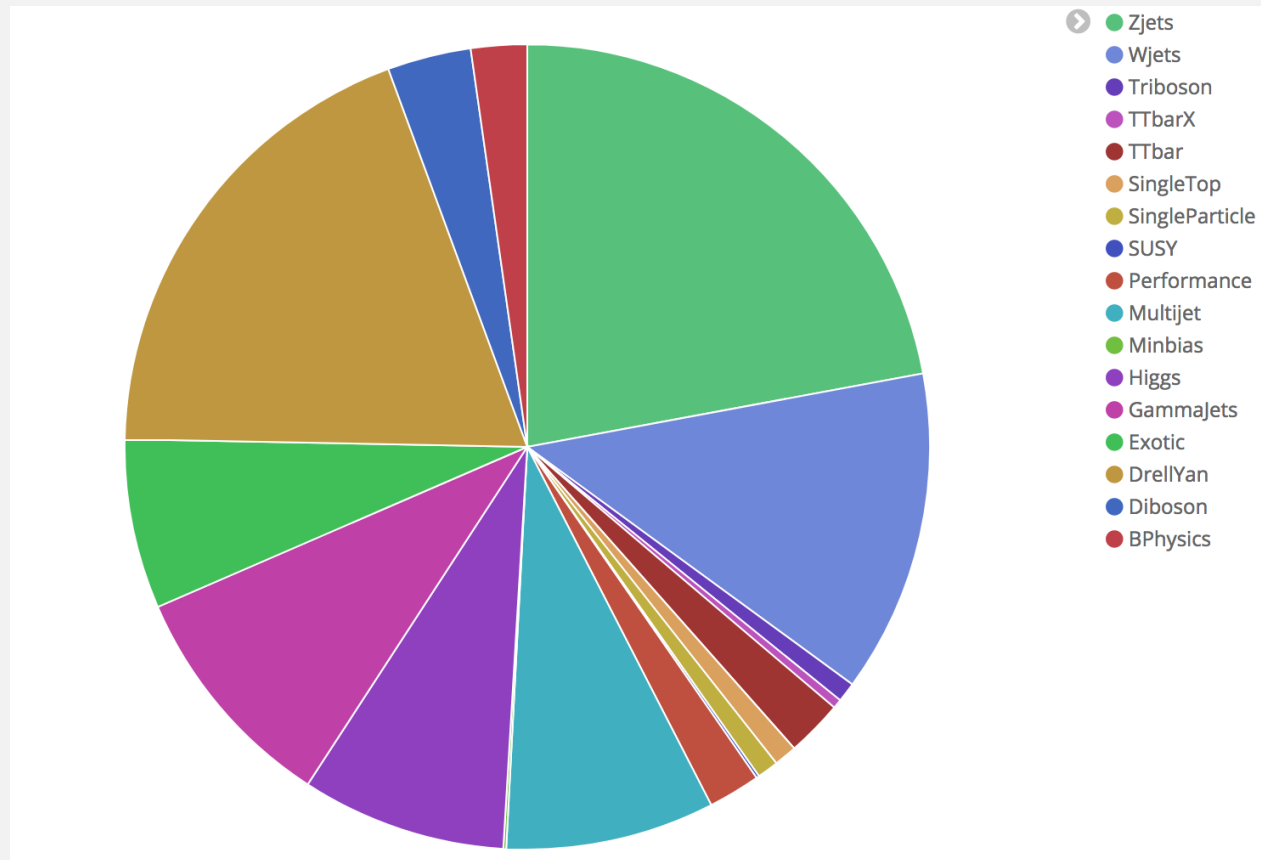
MACHINE LEARNING ISSUE FOR THE DKB PROJECT

Maria Grigorieva

THE SCOPE

- Define hashtags & physics categories for each job option
 - **Physics Short:** MadGraphPythia8EvtGen_AI4NNPDF23LO_MCPs
 - **Hashtags:** "0jet, BSM, MCI6b_TRIG, MadGraphPythia8EvtGen, drellYan, exotic"
 - **Physics Categories:** DrellYan, Exotic

INITIAL METADATA. PHYSICS CATEGORIES



INITIAL METADATA. HASHTAGS

Physics Short	Hashtag list
Pythia8BEvtGen_A14_CTEQ6L1_Bplus_Jpsi_mu3p5mu3p5_Kplus	2muon, BPhysics, Bplus, Jpsi, MCI6b_TRIG, Pythia8BEvtGen, exclusive
Pythia8BPhotospp_A14_CTEQ6L1_Bs_Jpsimu3p5mu3p5_phi	2muon, BPhysics, Bs, Jpsi, MCI6b_TRIG, Pythia8BPhotospp, exclusive
Pythia8BPhotospp_A14_CTEQ6L1_Bs_Jpsimu3p5mu3p5_phi	2muon, BPhysics, Bs, Jpsi, MCI6b_TRIG, Pythia8BPhotospp, exclusive
Pythia8B_A14_CTEQ6L1_Bs_mu3p5mu3p5	2muon, BPhysics, Bs, MCI6b_TRIG, Pythia8B, exclusive, rareDecay
Pythia8BEvtGen_A14_CTEQ6L1_Bs_mu3p5mu3p5	2muon, BPhysics, Bs, MCI6b_TRIG, Pythia8BEvtGen, exclusive, rareDecay
Pythia8BEvtGen_A14_CTEQ6L1_Bs_mu3mu3	2muon, BPhysics, Bs, MCI6b_TRIG, Pythia8BEvtGen, exclusive, rareDecay
ParticleGunEvtGen_Jpsi_mu5p5mu5p5_highd0	2muon, BPhysics, Jpsi, MCI6b_TRIG, ParticleGunEvtGen
Pythia8BPhotospp_A14_CTEQ6L1_bb_Jpsimu5p5mu5p5	2muon, BPhysics, Jpsi, MCI6b_TRIG, Pythia8BPhotospp, bottom, inclusive
Pythia8BPhotospp_A14_CTEQ6L1_bb_Jpsimu2p5mu2p5	2muon, BPhysics, Jpsi, MCI6b_TRIG, Pythia8BPhotospp, bottom, inclusive

INITIAL METADATA.VOLUMES

- Number of all physics shorts in ProdSys2 ~183 000 records
- Hashtags are defined for ~16 000 records
- Physics Categories could be defined explicitly for only 8% of all records

HASHTAG - PHYSICS CATEGORIES MAPPING

```
PHYS_CATEGORIES_MAP = {"BPhysics":["charmonium","Jpsi","Bs","Bd","Bminus","Bplus","CHARM','BOTTOM','BOTTOMONIUM','B0'],
    "BTag":["bTagging", "btagging"],
    "Diboson":["diboson","ZZ", "WW", "WZ", "WWbb", "WWll", "zz", "ww", "wz", "wwbb","wll"],
    "DrellYan":["drellyan", "dy"],
    "Exotic":["exotic", "monojet", "blackhole", "technicolor", "RandallSundrum",
        "Wprime", "Zprime", "magneticMonopole", "extraDimensions", "warpedED",
        "randallsundrum", "wprime", "zprime", "magneticmonopole",
        "extradimensions", "warpeded", "contactInteraction","contactinteraction",'SEESAW'],
    "GammaJets":["photon", "diphoton"],
    "Higgs":["WHiggs", "ZHiggs", "mH125", "Higgs", "VBF", "SMHiggs", "higgs", "mh125",
        "zhiggs", "whiggs", "bsmhiggs", "chargedHiggs","BSMHiggs","smhiggs"],
    "Minbias":["minBias", "minbias"],
    "Multijet":["dijet", "multijet", "qcd"],
    "Performance":["performance"],
    "SingleParticle":["singleparticle"],
    "SingleTop":["singleTop", "singletop"],
    "SUSY":["SUSY", "pMSSM", "leptoSUSY", "RPV", "bino", "susy", "pmssm", "leptosusy", "rpv","MSSM"],
    "Triboson":["tripleGaugeCoupling", "triboson", "ZZW", "WWW", "triplegaugecoupling", "zzw", "www"],
    "TTbar":["ttbar"],
    "TTbarX":["ttw","ttz","ttv","ttvv","4top","ttW","ttZ","ttV","ttWW","ttVV"],
    "Upgrade":["upgrad"],
    "Wjets":["W", "w"],
    "Zjets":["Z", "z"]}

if 'singletop' in phys_short: return "SingleTop"
if 'ttbar' in phys_short: return "TTbar"
if 'jets' in phys_short: return "Multijet"
if 'h125' in phys_short: return "Higgs"
if 'ttbb' in phys_short: return "TTbarX"
if 'ttgamma' in phys_short: return "TTbarX"
if '_tt_' in phys_short: return "TTbar"
if 'upsilon' in phys_short: return "BPhysics"
if 'tanb' in phys_short: return "SUSY"
if '4topci' in phys_short: return "Exotic"
if 'xhh' in phys_short: return "Higgs"
if '3top' in phys_short: return "TTbarX"
if '_wt' in phys_short: return "SingleTop"
if '_wwbb' in phys_short: return "SingleTop"
if '_wenu_' in phys_short: return "Wjets"
```

TASK

- Use machine learning algorithms to propagate hashtags to all job options in ProdSys2
 - Define physics category for each record, where possible (using hashtag-category mapping)
 - Use 16 000 records with hashtags to implement training set
 - Apply training set to all records
 - Result:
 - Table with [*phys_short | hashtags | physics categories*](#) for all records