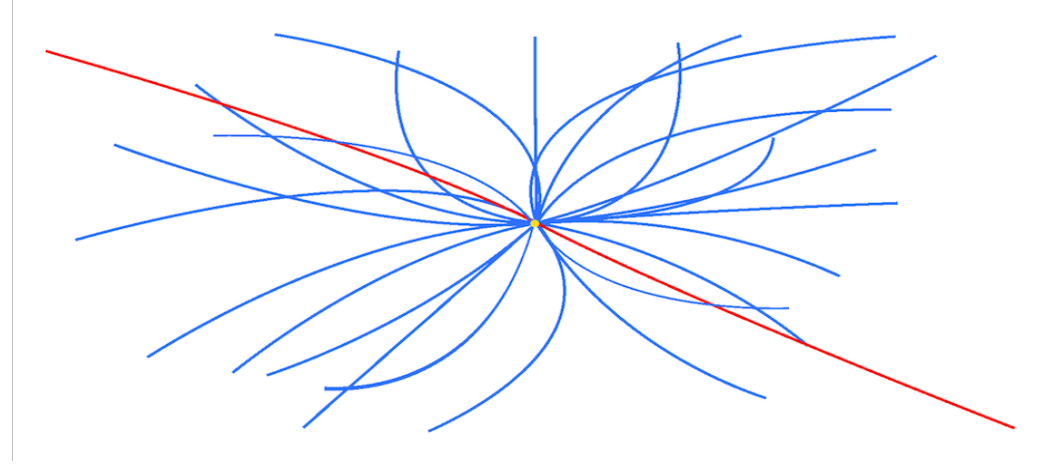


PFVA'25 Kış Okulu Ders Notları



High Energy Physics data for everyone

Açık Veri ile Uygulama I & II

A.C. Canbay, I.T. Turk Cakir, O. Cakir
(AU)



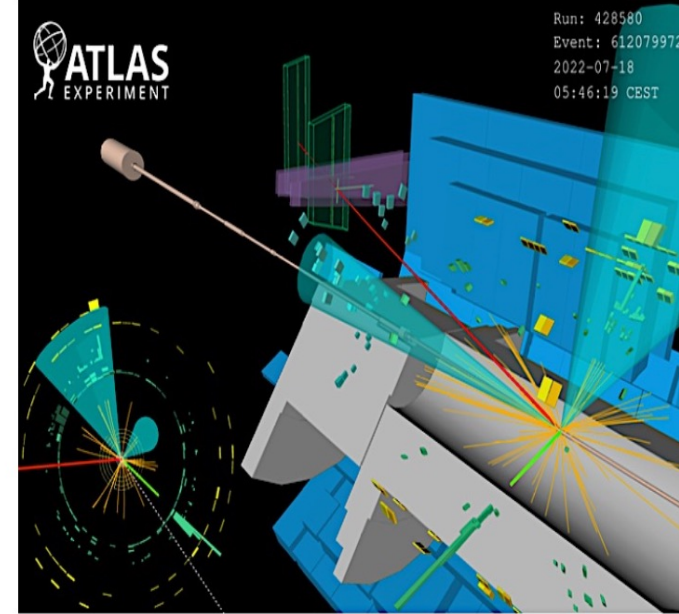
ATLAS DENEYİ



ATLAS deneyi, parçacık fiziğinin standart modeli (SM)'in tanımladığı süreçleri hassas incelemektedir. Tesir kesiti ölçümleri, etkileşme bağlaşımlarının incelenmesi, kinematik değişkenlerin hassas belirlenmesi, parçacık tanıma, yeniden oluşturma (reconstruction), sinyal ve ardaalan olaylarının incelenmesi, olayların veri analizinde ileri yöntemleri uygulayan çalışmalar/araştırmalar yapmaktadır.

Bununla birlikte, standart model ötesi (BSM) olası beklenen yeni fiziği (yeni parçacıklar, yeni etkileşmeler, karanlık madde, vd.) araştırmaktadır.

Evrenin gizemini araştırmak



pp → tt → evμν+jets Run-3 candidate



Open Data



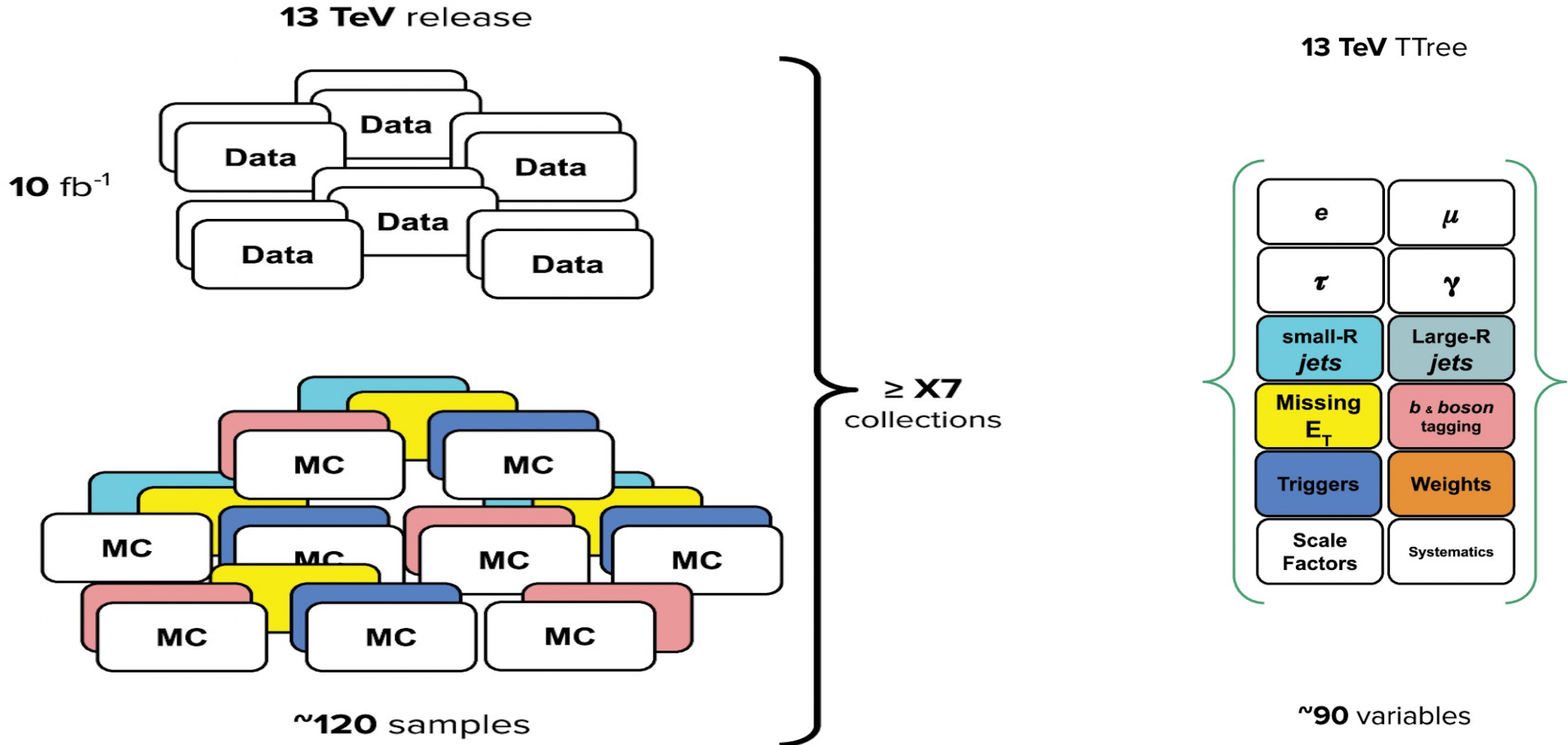
The screenshot shows the ATLAS Open Data website interface. At the top left is the 'opendata CERN' logo. To the right is a navigation menu with links for 'ABOUT', 'SEARCH', 'EDUCATION', and 'RESEARCH'. The main content area is split into two columns. The left column is titled 'Education' and contains the text 'Visualise events, check reconstructed data, run tools or build your own!' and a 'Start learning' button. The right column is titled 'Research' and contains the text 'Get the genuine working environments, virtual machines and datasets to start your research' and a 'Start analysing' button. In the center, there is a diagram of particle tracks with several nodes labeled with Greek letters: μ (muon), γ (photon), τ (tau lepton), q (quark), and e (electron). A diagonal cyan line runs across the diagram.

Open Data

- Fizik analizleri, milyonlarca satır C++ kodunu ve Python çalıştırmak için yüzlerce petabayt veri depolama alanı ve yüz binlerce CPU gerektirir.
- ATLAS Open Data
 - öğrenciler, fizikçiler ve hatta fizikçi olmayanlar için hazırlanmış açık verileri içeriyor. <https://opendata.atlas.cern/>



Open Data @ 13 TeV



Open Data

- Open Data, eğitim, kurs (tutorial), araştırma vb. öğrenme, öğretme ve araştırma etkinliklerinde kullanılabilir.
- Datasetler, MC ve Data için tanımlanmıştır, DAOD->Ntuple/Mini olay veri formatında 7/8 TeV ve 13 TeV enerjilerinde proton-proton çarpışması kısmi (hepsi değil, çok az bir kesri bulunuyor) verilerini içermektedir.
- Analize yeni başlayanlar için, analiz kodu geliştirenler için, ML uygulayanlar için, vb. kullanışlı. Deneylerin Fizik Analiz gruplarında bir analiz yapmak için bir ön çalışma olarak kullanılabilir.
- Tek başına veya birkaç kişi makale yazması ve yayınlaması için uygun değil / izin verilmiyor.
- Deneylerin yazar (authorship) olma koşulları ve MC ve Data kullanımı koşulları bulunuyor, bunlara uymak gerekir.

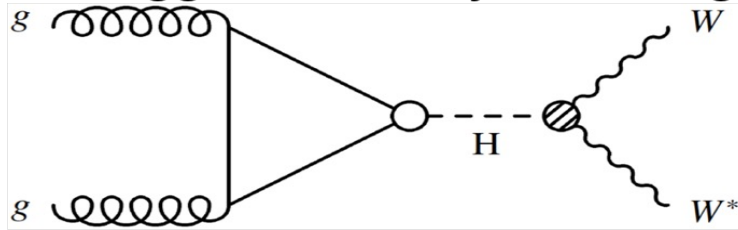
* Team CERN Open Data lisansı altında serbest kullanılabilir.

ATLAS_OpenData_Example 1: pp->H->WW (signal), pp->WW (background)

Notebook example: simple two-samples comparison

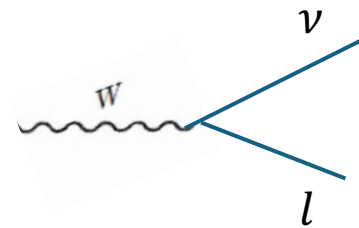
The following analysis is comparing the kinematics between events coming from the SM Higgs boson decaying to 2 W-bosons to those coming from the SM WW-diboson background production.

SM Higgs to WW Feynman diagram:

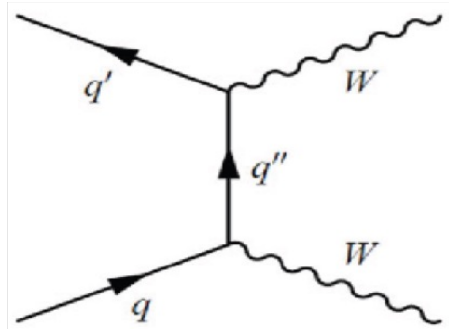


Signal

Bozunma (decay)



SM WW-diboson Feynman diagram:



Background

```
[1]: import ROOT
import numpy as np
```

```
[2]: bkg = ROOT.RDataFrame("mini", "mc_363492.llvv.2lep.root")
bkg = (
    bkg.Define("met_et0", "met_et/1000")
    .Define("lep_pt0", "lep_pt[0]/1000")
    .Define("mT", "sqrt( 2*lep_pt[0]*met_et*(1 - cos(lep_phi-met_phi)) )/1000")
)

sig = ROOT.RDataFrame("mini", "mc_345324.ggH125_WW2lep.2lep.root")
sig = (
    sig.Define("met_et0", "met_et/1000")
    .Define("lep_pt0", "lep_pt[0]/1000")
    .Define("mT", "sqrt( 2*lep_pt[0]*met_et*(1 - cos(lep_phi-met_phi)) )/1000")
)

c = ROOT.TCanvas("testCanvas", "a first way to plot a variable", 800, 600)
c.Divide(2, 2)
ROOT.gStyle.SetOptStat("")

c.cd(1)

hist_sig = sig.Histo1D(("h_met_et", "", 40, 0, 200), "met_et0")
hist_sig.Scale(1/hist_sig.Integral())
hist_sig.SetLineColor(2)

hist_sig.SetXTitle("Transverse mass [GeV]")
hist_sig.SetYTitle("Number of leptons / 5 GeV")
hist_sig.Draw("hist")
```

```
hist_bkg = bkg.Histo1D(("h_met_et", "", 40, 0, 200), "met_et0")
hist_bkg.Scale(1/hist_bkg.Integral())
hist_bkg.SetLineColor(4)
hist_bkg.Draw("hist,same")
```

```
c.cd(2)
```

```
hist2_sig = sig.Histo1D(("h_lep_pt", "", 40, 0, 200), "lep_pt0")
hist2_sig.Scale(1/hist2_sig.Integral())
hist2_sig.SetLineColor(2)
```

Lepton pT

```
hist2_sig.SetXTitle("Transverse mass [GeV]")
hist2_sig.SetYTitle("Number of leptons / 5 GeV")
hist2_sig.Draw("hist")
```

```
hist2_bkg = bkg.Histo1D(("h_lep_pt", "", 40, 0, 200), "lep_pt0")
hist2_bkg.Scale(1/hist2_bkg.Integral())
hist2_bkg.SetLineColor(4)
hist2_bkg.Draw("hist,same")
```

```
c.cd(3)
```

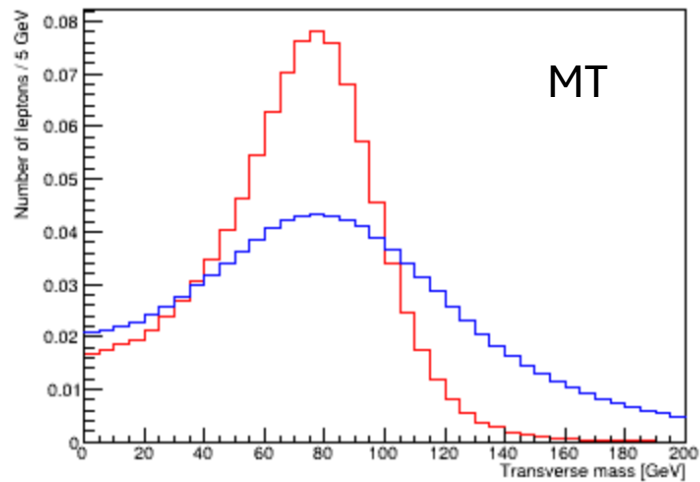
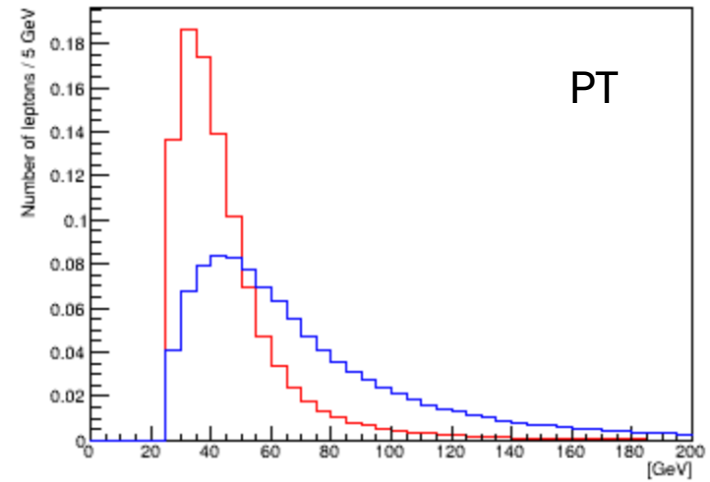
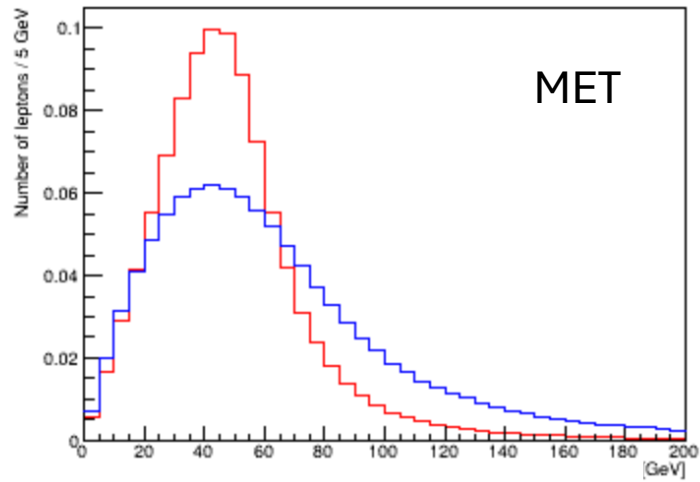
```
hist3_sig = sig.Histo1D(("h_mT", "", 40, 0, 200), "mT")
hist3_sig.Scale(1/hist3_sig.Integral())
hist3_sig.SetLineColor(2)
```

```
hist3_sig.SetXTitle("Transverse mass [GeV]")
hist3_sig.SetYTitle("Number of leptons / 5 GeV")
hist3_sig.Draw("hist")
```

```
hist3_bkg = bkg.Histo1D(("h_mT", "", 40, 0, 200), "mT")
hist3_bkg.Scale(1/hist3_bkg.Integral())
hist3_bkg.SetLineColor(4)
hist3_bkg.Draw("hist,same")
```

```
c.Draw()
c.Update()
```

Çıktılar





ATLAS OPEN DATA UYGULAMA II

(H->ZZ->4l)

A.C. Canbay, I.T. Turk Cakir, O. Cakir

(AU)

Setup and Run Steps

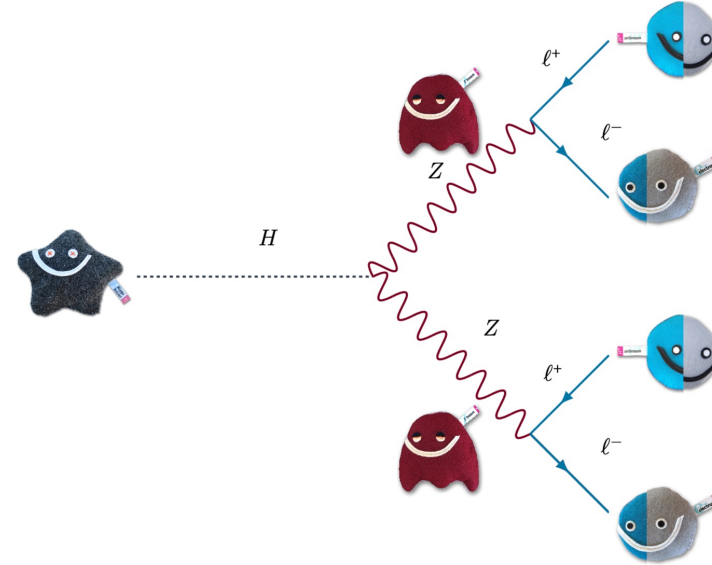
Bir Jupyter not defteri çalıştırma
Bilgisayarınızda ilk kez kurulum için bazı ayarlar (mybinder'da gerekmez)
Her sefer çalıştırmada yapılacaklar
Lumi, oran, dosya yolu
Örneklemler (samples)
Bir kesme değiştirme
Bir kesme uygulama
Optimizasyon
Güçlendirilmiş Karar Ağacı (BDT)
Bölmeleme, eğitime ve test etme
Karar Ağaçlarını eğitime
Bir Sınıflandırıcının Performansını değerlendirme
Alıcı İşletim Karakteristiği (ROC) eğrisi
Aşırı eğitim kontrolü
Optimizasyon
Daha ileriye...

https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/32943340fa1aada639c96f10a2224bd671583cca/13-TeV-examples/uproot_python/#BDT

Search for $H \rightarrow ZZ$ with BDT

Bu python not defteri (.ipynb), Higgs bozonunu aramada Makine Öğrenimini uygulama adımlarını göstermek için ATLAS Açık Veri'yi kullanır.
<http://opendata.atlas.cern>

ATLAS Açık Veri, eğitim amaçlı olarak LHC'deki proton-proton çarpışma verilerine açık erişim sağlar. ATLAS Açık Veri kaynakları lise, üniversite lisans ve lisansüstü öğrenciler için idealdir.



** Not defterleri (notebooks), işiniz için Makine Öğrenimini geliştirmek için mükemmel bir platformdur, çünkü tam olarak şunlara ihtiyacınız olacak: kod, görselleştirmeler ve anlatı metni!*

Search for H->ZZ with BDT

Bu not defteri sinyal/arka plan sınıflandırmasına odaklanacaktır. Bu not defterinin sonunda şunları yapabileceksiniz:

[1] sinyal ve arka planı sınıflandırmak için Güçlendirilmiş Karar Ağacı (BDT) çalıştırma

[2] Güçlendirilmiş Karar Ağacınızı geliştirmek için değiştirebileceğiniz bazı parametrelerin bilinmesi

Bize yardımcı olması için bir dizi araç kullanacağız:

uproot: Parçacık fiziğinde genellikle kullanılan .root dosyalarını Makine Öğrenmesinde kullanılan veri biçimlerine okumamızı sağlar

pandas: Verileri Makine Öğrenmesinde yaygın olarak kullanılan bir biçim olan veri çerçeveleri olarak depolamamızı sağlar

numpy: Histogramlama gibi sayısal hesaplamalar sağlar

matplotlib: Grafikler, şekiller, resimler, görselleştirmeler oluşturmak için yaygın bir araç

Jupyter-lab ile Uygulama

Geliştirme...

Kodu daha geliştirmek isterseniz, deneyebileceğiniz birkaç şey var:

'Lumi, kesir, dosya yolu'nda işlenen olayların oranını artırın.

'Örneklemler'e diğer H->ZZ sinyal örneklerini ekleyin. Bunları example2.ipynb'den kopyalayabilirsiniz. Önce bunları tek tek eklemeyi deneyin, sonra eklenen her şeyle birlikte işlerin nasıl görüldüğüne bakın.

'Örneklemler'e Z ve ttbar arka plan örneklerini ekleyin. Bunları example2.ipynb'den kopyalayabilirsiniz. Önce bunları ayrı ayrı eklemeyi deneyin, sonra eklenen her şeyle birlikte işlerin nasıl görüldüğüne bakın.

'Güçlendirilmiş Karar Ağacı'na (BDT) birkaç değişken daha ekleyin. Hepsini birden eklemek yerine tek tek ekleyin, çünkü bir değişken eklemek, anti-korelasyon nedeniyle BDT performansını düşürebilir. Değişkenler hakkında bazı fikirler edinmek için ATLAS tarafından Higgs bozonunun keşfine ilişkin makaleye bakabilirsiniz.

'Örneklemler'e gerçek veriler ekleyin ve verilerdeki ve simülasyondaki BDT çıktı dağılımlarının eşleşip eşleşmediğine bakın. Verileri example2.ipynb'den kopyalayabilirsiniz.

'Karar Ağaçlarını Eğitme' bölümünde bazı BDT hiper parametrelerini değiştirin. Her değişiklikte şunlara dikkat edin: ROC eğrisinin altındaki toplam alan, BDT çıktı dağılımında sinyal ve arka plan arasındaki ayırım elde edilebilecek S/B veya S/\sqrt{B} bulun.

BDT'mizi simüle edilmiş veriler üzerinde eğittiğimizi ve test ettiğimizi fark edin. Daha sonra bunu gerçek deneysel verilere uygulanabilir. BDT'nizden memnun kaldığınızda, tüm verileri çalıştırmak için onu tam bir analize geri koymak mümkün olabilir.

References

[1] ATLAS Open Data <http://opendata.atlas.cern>

[2] Particle Zoo, <https://www.particlezoo.net/>

[3] PDG 2024