Study with new proposed features for various AE

Rishabh Uniyal
Catholic U, Washington DC
September 9, 2019

Histograms of quantiles per run

https://cernbox.cern.ch/index.php/apps/files/?dir=/__myshares/ML4DQM_histograms+%28id%3A201154%29

We need to chose relevant quantiles for hists because the number of histograms is very large and I get and os error

OSError: [Errno 27] File too large: '321760_qPFJetPhi_2_jetht.png'
So, could not produce for all quantiles in good data for single muon and zerobias

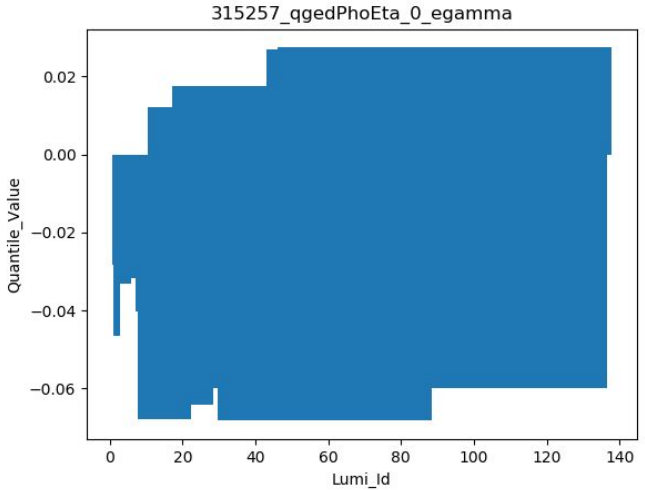
Code used to produce hists

https://github.com/rishabhCMS/ML4DQM-Visualizations/blob/master/DataViz2018/allplots.py

Questions

Negative values in some columns?

Egamma = Index(['qpVtxY__0', 'qpVtxY__2', 'qpVtxY__3', 'qpVtxY__4', 'qpVtxY__5',
'qpVtxY__6', 'qpVtxZ__0', 'qpVtxZ__2', 'qpVtxZ__3', 'qpVtxZ__4',
'qpVtxZ__5', 'qpVtxZ__6', 'qGsfEta_0', 'qGsfEta_2', 'qGsfEta_3',
'qGsfEta_4', 'qGsfPhi_0', 'qGsfPhi_2', 'qGsfPhi_3', 'qGsfPhi_4',
'qgedPhoEta_0', 'qgedPhoEta_2', 'qgedPhoEta_3', 'qgedPhoEta_4',
'qgedPhoPhi_0', 'qgedPhoPhi_2', 'qgedPhoPhi_3', 'qgedPhoPhi_4'],
dtype='object')



PD Features mapping

new features (electron in EGamma)
Features removed

ZeroBias	JetHT	EGamma	SingleMuon
qpVtxChi2 qpVtxNtr qpVtxX qpVtxY qpVtxZ qPUEvt qlumiEvt qgTkPt qgTkEta qgTkPhi qgTkN qgTkChi2 qgTkNHits qgTkNLay 7*13 = 91 features	qpVtxChi2 qpVtxNtr qpVtxX qpVtxY qpVtxZ qPUEvt qlumiEvt PFJetN PFJetPt PFJetPhi PFJetEta PFMetPt PFMetPhi CalJetN CalJetPt CalJetEta CalJetPhi CalJetEn CalMETPt CalMETPhi CCEn_ CCEta_ CCPhi_ SCEn_ SCEta_ SCPhi_ 7*24 = 168	qpVtxChi2 qpVtxNtr qpVtxX_ qpVtxY_ qpVtxZ_ qPUEvt_ qlumiEvt **qGsfPt** **qGsfEta** **qGsfPhi** **qGsfN** qPhoN gedPhoPt gedPhoEta gedPhoPhi gedPhoEn gedPhoe1x5 gedPhoe3x3 **SigmaIEta_; SigmaIPhi_; r9_; HadOEm_; drSumPt_; drSumEt_; eSCOP_; ecEn_;** 7*30 = 210	qpVtxChi2_ qpVtxNtr_ qpVtxX_ qpVtxY_ qpVtxZ_ qPUEvt_ qlumiEvt_ qglobTkN qglobTkPt qglobTkEta qglobTkPhi _ qglobTkChi2 qglobTkNHits **qMuNCh** qMuN qMuPt qMuEta qMuPhi qMuEn qMuChi2 7*18 = 126

4

Features from the twiki

ZeroBias	JetHT	EGamma	SingleMuon
qpVtxChi2	qpVtxChi2	qpVtxChi2	qPUEvt
qpVtxNtr	qpVtxNtr	qpVtxNtr	qlumiEvt
qPUEvt	qPUEvt		
qlumiEvt	qlumiEvt	qPUEvt	qMuN
		qlumiEvt	qMuNCh
qgTkPt	qPFJetN		qMuPt
qgTkEta	qPFJetPt	qGsfPt	qMuEta
qgTkPhi	qPFJetPhi	qGsfEta	qMuPhi
qgTkN	qPFJetEta	qGsfPhi	qMuEn
qgTkChi2	qPFMetPt	qGsfN	qMuChi2
qgTkNHits	qPFMetPhi		
qgTkNLay		qPhoN	
	qCalJetN	gedPhoPt	
	qCalJetPt	gedPhoEta	
	qCalJetEta	gedPhoPhi	
	qCalJetPhi	gedPhoEn	
	qCalJetEn	gedPhoe3x3	
	qCalMETPt		
	qCalMETPhi	qSigmaIEta	
		qSigmaIPhi	
	qCCEn		
	qCCEta		
	qCCPhi		
	qSCEn		
	qSCEta		
	qSCPhi		

PD Features Mapping From the twiki

new features
(electron in EGamma)
Features removed

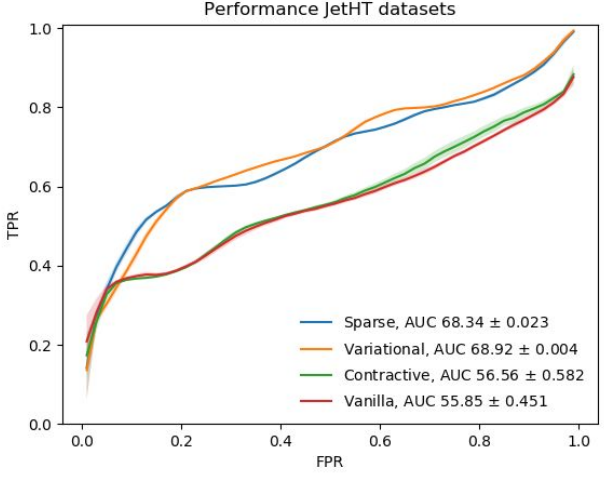
ZeroBias	JetHT	EGamma	SingleMuon
qpVtxChi2 qpVtxNtr qpVtxX qpVtxY qpVtxZ qPUEvt qlumiEvt qgTkPt qgTkEta qgTkPhi qgTkN qgTkChi2 qgTkNHits qgTkNLay 11*13 = 143 features	qpVtxChi2 qpVtxNtr qpVtxX qpVtxY qpVtxZ qPUEvt qlumiEvt qPFJetN qPFJetPt qPFJetPhi qPFJetEta qPFMetPt qPFMetPhi qCalJetN qCalJetPt qCalJetEta qCalJetPhi qCalJetEn qCalMETPt qCalMETPhi qCCEn_ qCCEta_ qCCPhi_ qSCEn_ qSCEta_ qSCPhi_ 7*23 = 161	qpVtxChi2 qpVtxNtr qpVtxX_ qpVtxY_ qpVtxZ_ qPUEvt_ qlumiEvt **qGsfPt** **qGsfEta** **qGsfPhi** **qGsfN** qPhoN gedPhoPt gedPhoEta gedPhoPhi gedPhoEn gedPhoe1x5 gedPhoe3x3 **qSigmaIEta_;** **qSigmaIPhi_; r9_;** **HadOEm_; drSumPt_;** **drSumEt_; eSCOP_;** **ecEn_;** 7*16 = 112	qpVtxChi2_ qpVtxNtr_ qpVtxX_ qpVtxY_ qpVtxZ_ qPUEvt_ qlumiEvt_ qglobTkN qglobTkPt qglobTkEta qglobTkPhi _ qglobTkChi2 qglobTkNHits **qMuNCh** qMuN qMuPt qMuEta qMuPhi qMuEn qMuChi2 7*8 = 56

6

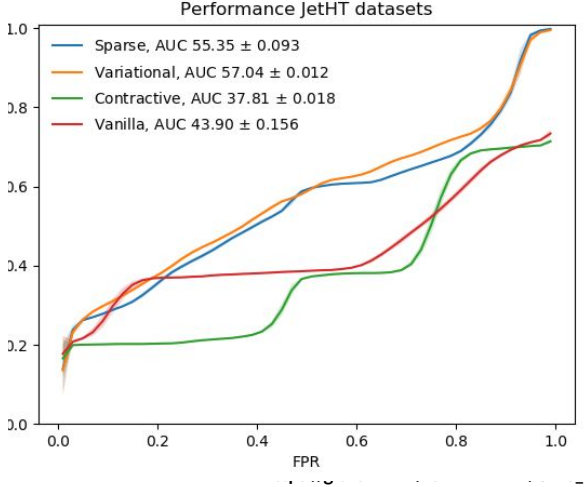
JetHT

of Epochs - 1200
BS = 2e15
Dataset 2018
# of bins 50

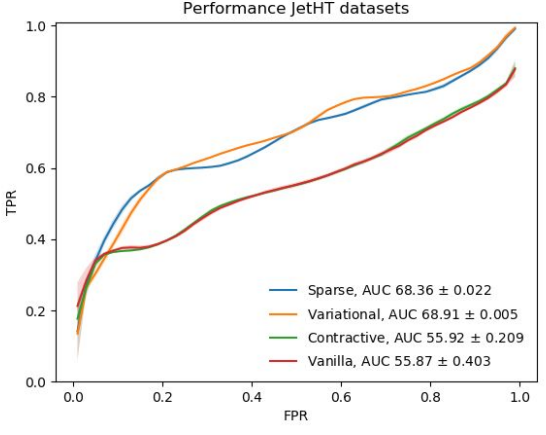
With changed features from slide 4



Previously used features by Jab



Twiki Features

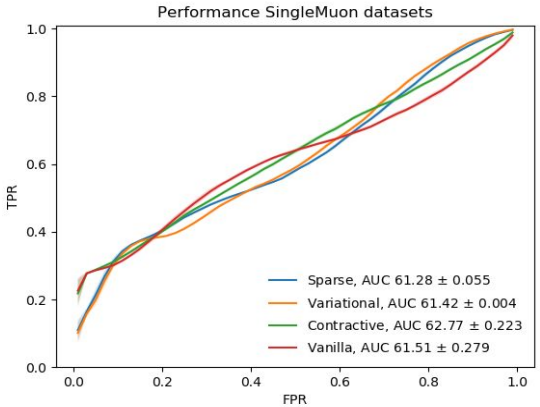


OMALY/blob/master/report/reco/new_data/reports/22july2019.pdf

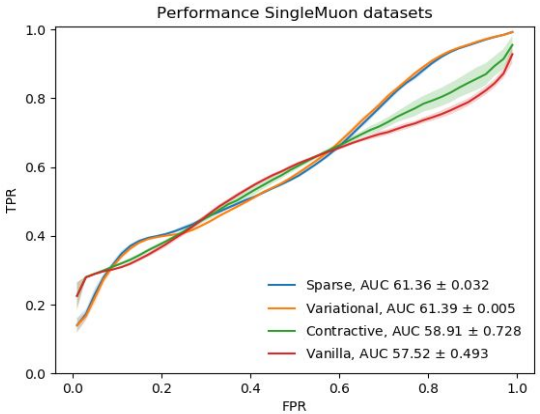
SingleMuon

of Epochs - 1200
BS = 2e15
Dataset 2018

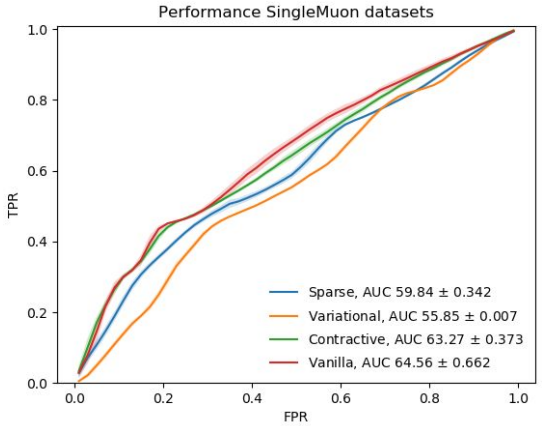
With changed features from slide 4



Previously used features by Jab



Twiki Features

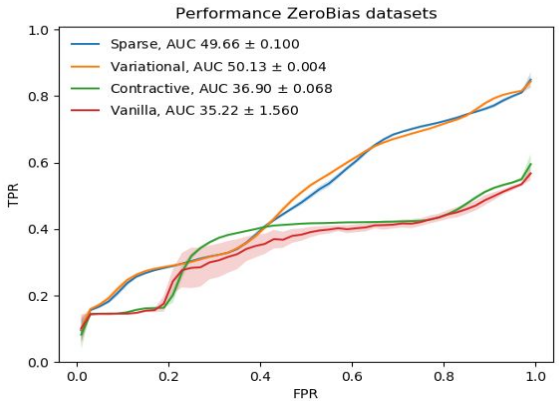


https://github.com/calzonelover/CMS_DC_ANOMALY/blob/master/report/reco/new_data/reports/22july2019.pdf

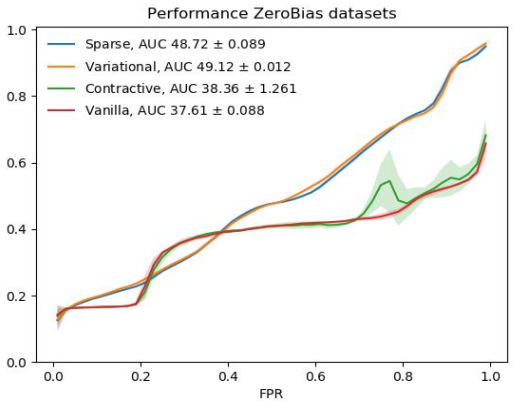
ZeroBias

of Epochs - 1200
BS = 2e15
Dataset 2018

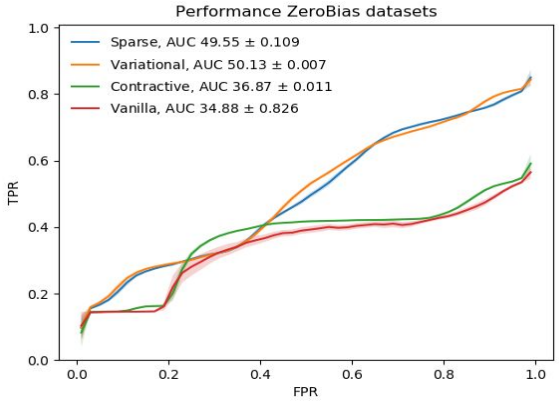
With changed features from slide 4



Previously used features by Jab



Twiki Features

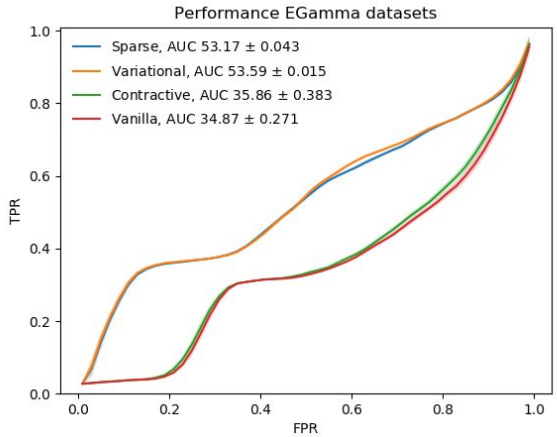


https://github.com/calzonelover/CMS_DC_ANOMALY/blob/master/report/reco/new_data/reports/22july2019.pdf

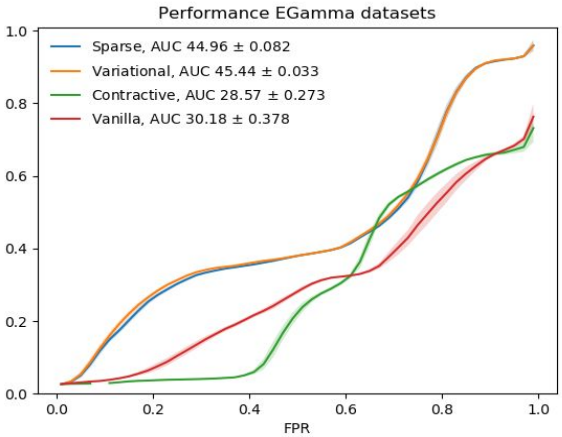
EGamma

of Epochs - 1200
BS = 2e15
Dataset 2018

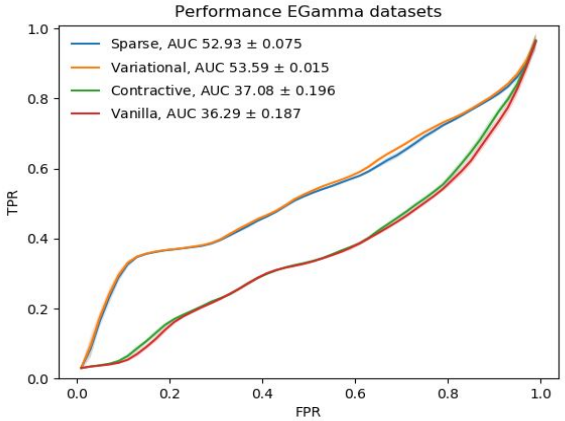
With changed features from slide 4



Previously used features by Jab



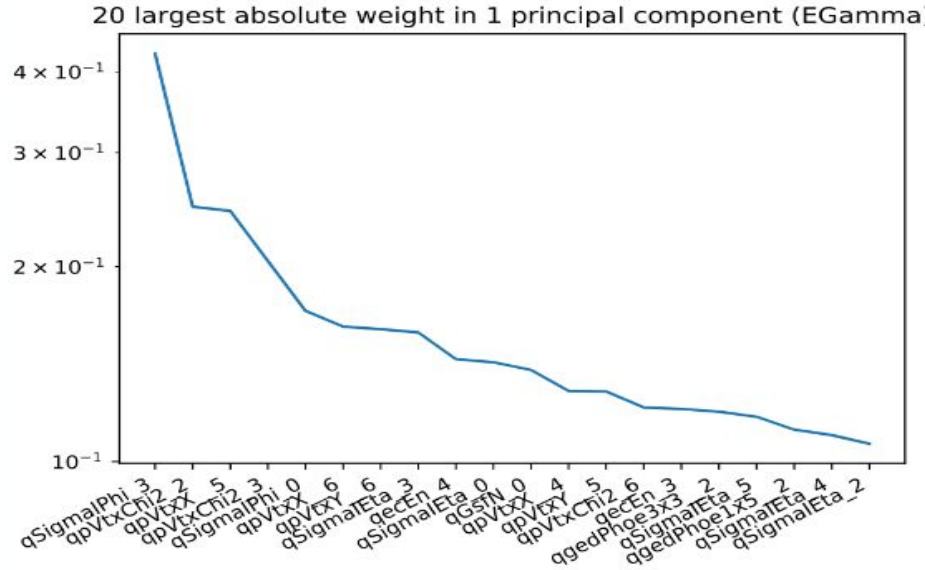
Twiki Features

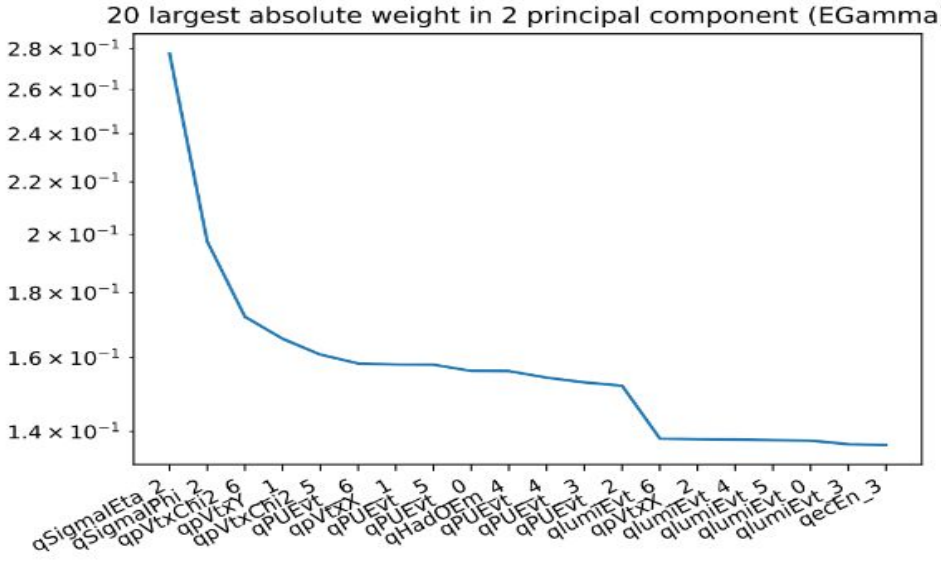


https://github.com/calzonelover/CMS_DC_ANOMALY/blob/master/report/reco/new_data/reports/22july2019.pdf

Backup

Why AE performs better with new features for EGamma? (PCA analysis by Jab)





qpVtx and qpVtxY are the dominant features so, removing them should decrease the performance of the AE

Dominated features
- **qpVtxX and qpVtxY**
- **qSigmaIEta**

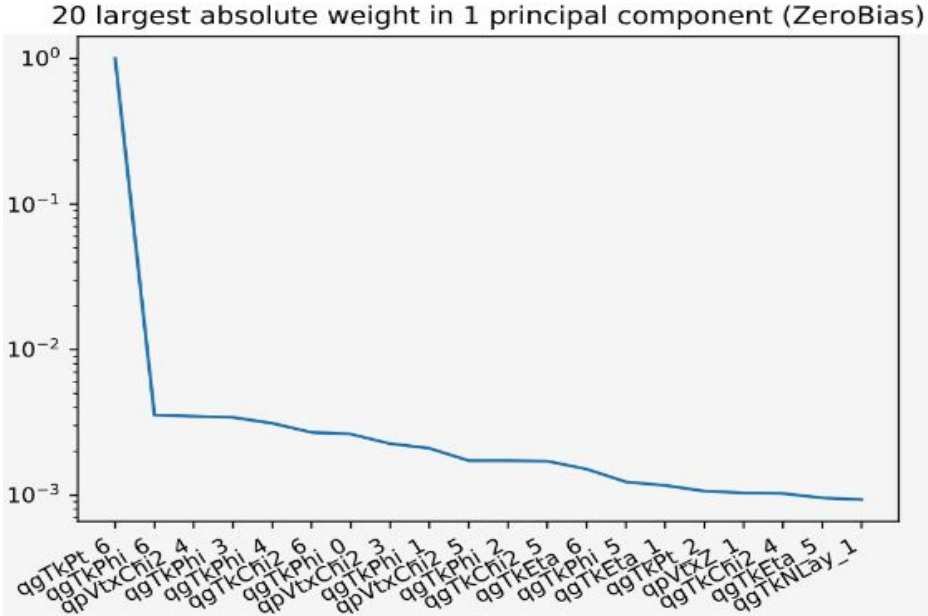
Overlapping feature
- **qSigmaIPhi**
- **qpVtxChi2**

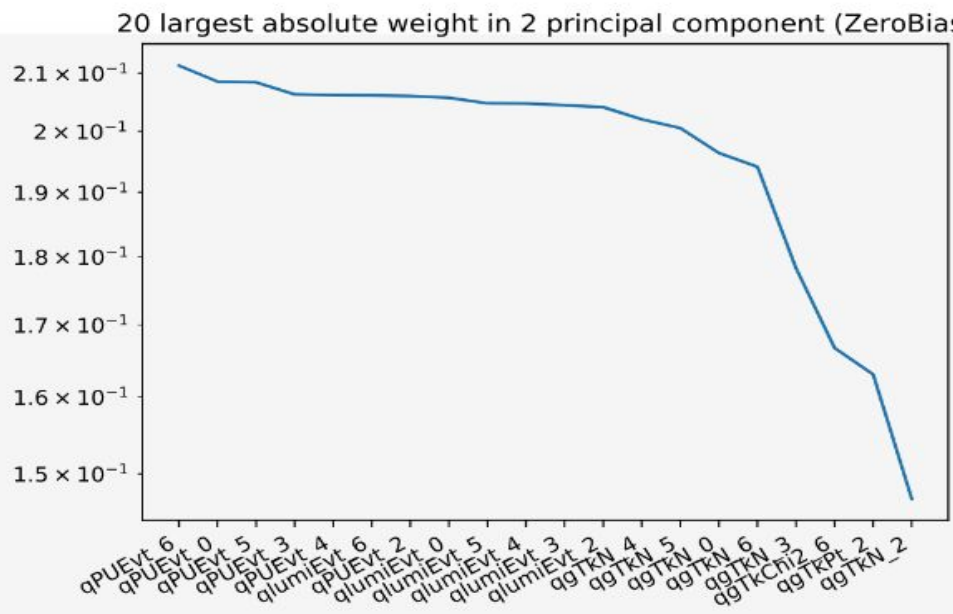
Dominated features
- **qPUEvt**
- **qlumiEvt**

Explained variance ratio ~ [0.31 0.25]

https://github.com/calzonelover/CMS_DC_ANOMALY/blob/master/report/reco/new_data/reports/12Aug2019.pdf

ZeroBias



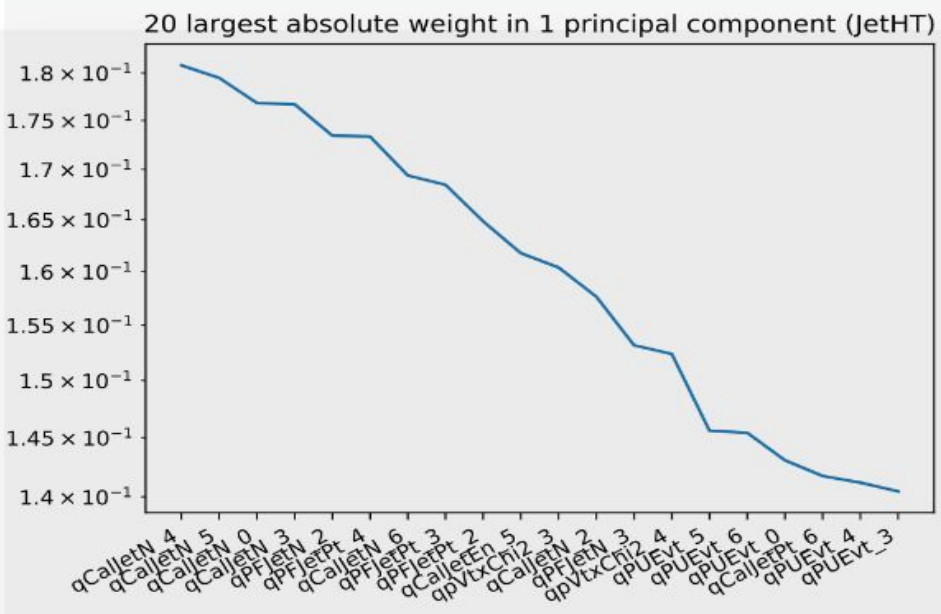


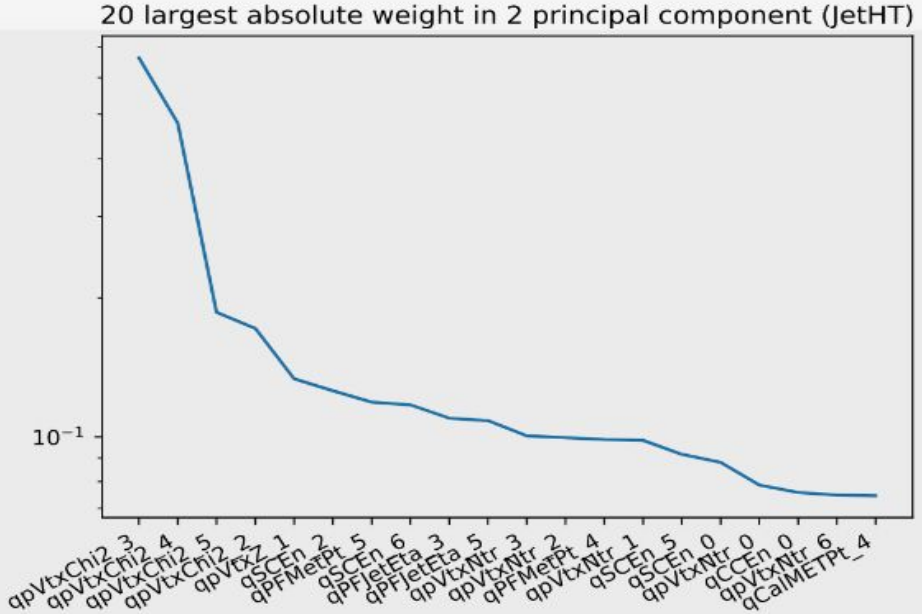
Dominated features
- **qgTkPt**
- **qgTkPhi**

Dominated features
- **qPUEvt**
- **qlumiEvt**
- **qgTkN**

https://github.com/calzonelover/CMS_DC_ANOMALY/blob/master/report/reco/new_data/reports/12Aug2019.pdf

JetHT

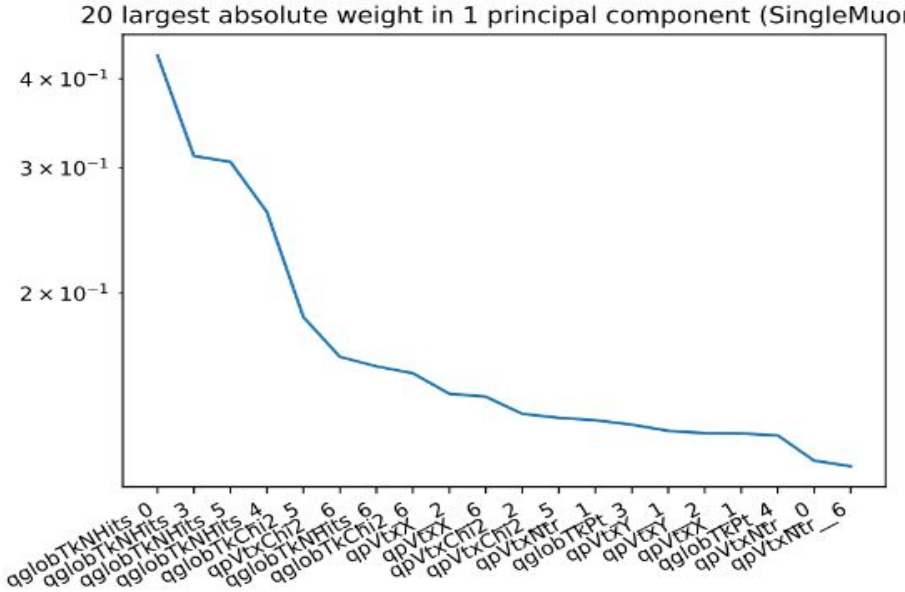


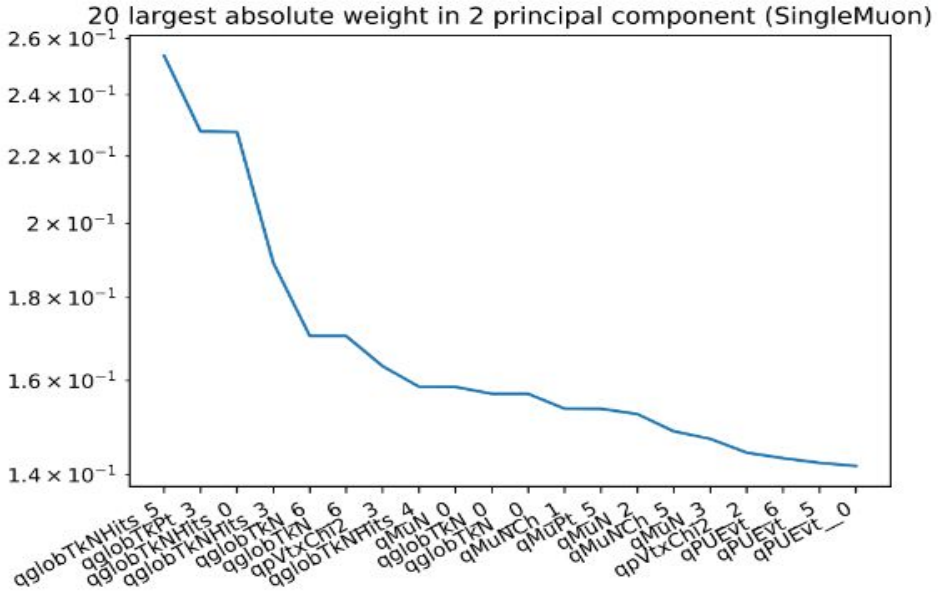


Dominated features
- **qCalJetN**
- **qCalJetPt**
- **qPUEvt**

Dominated features
- **qpVtxChi2**
- **qPFMetPt and qPFJetEta**

SingleMuon





Dominated features
- **qglobTkChi2**
- **qpVtxX and qpVtxY**

Overlapping feature
- **qglobTkNHits**

Dominated features
- **qPUEvt**
- **qMuN and qMuNCh**

Next steps??