

INTEGRATING ML ALGORITHMS FOR LHC DATA COMPRESSION INTO THE ESCAPE VIRTUAL RESEARCH ENVIRONMENT

Mentor: Caterina Doglioni

LARGE HADRON COLLIDER DETECTORS

1 billion particle collisions per second



generate about one petabyte of data



storage capacity limits the amount of information



we need to compress the data

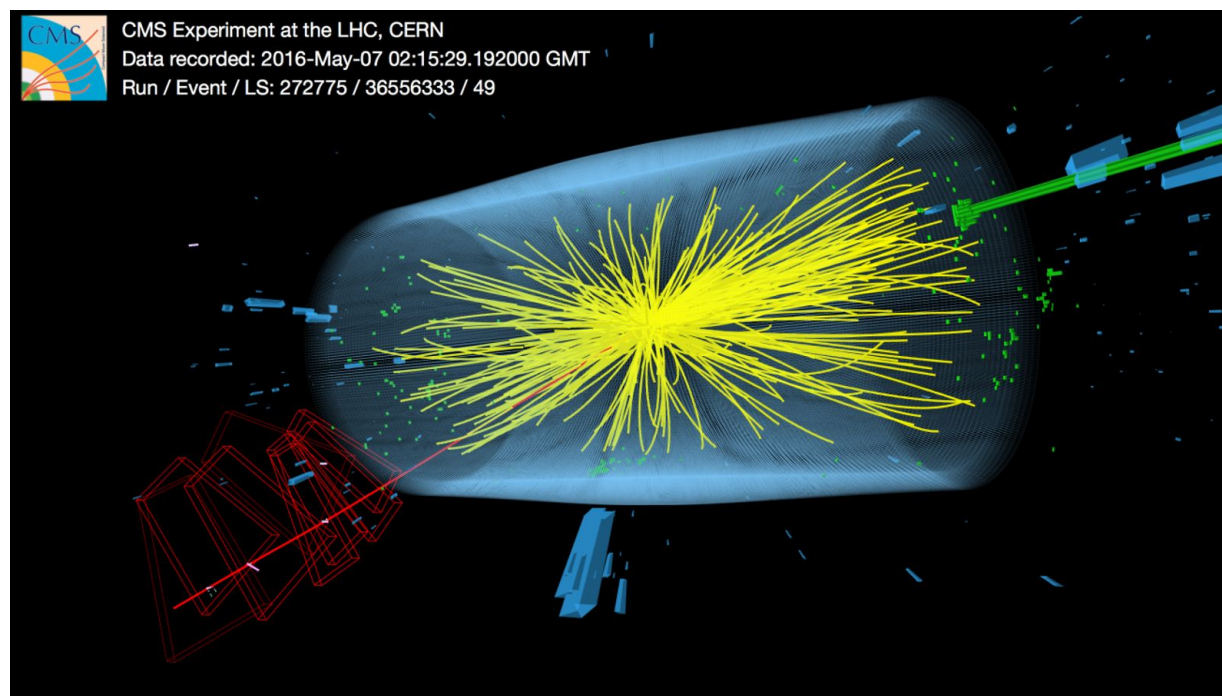
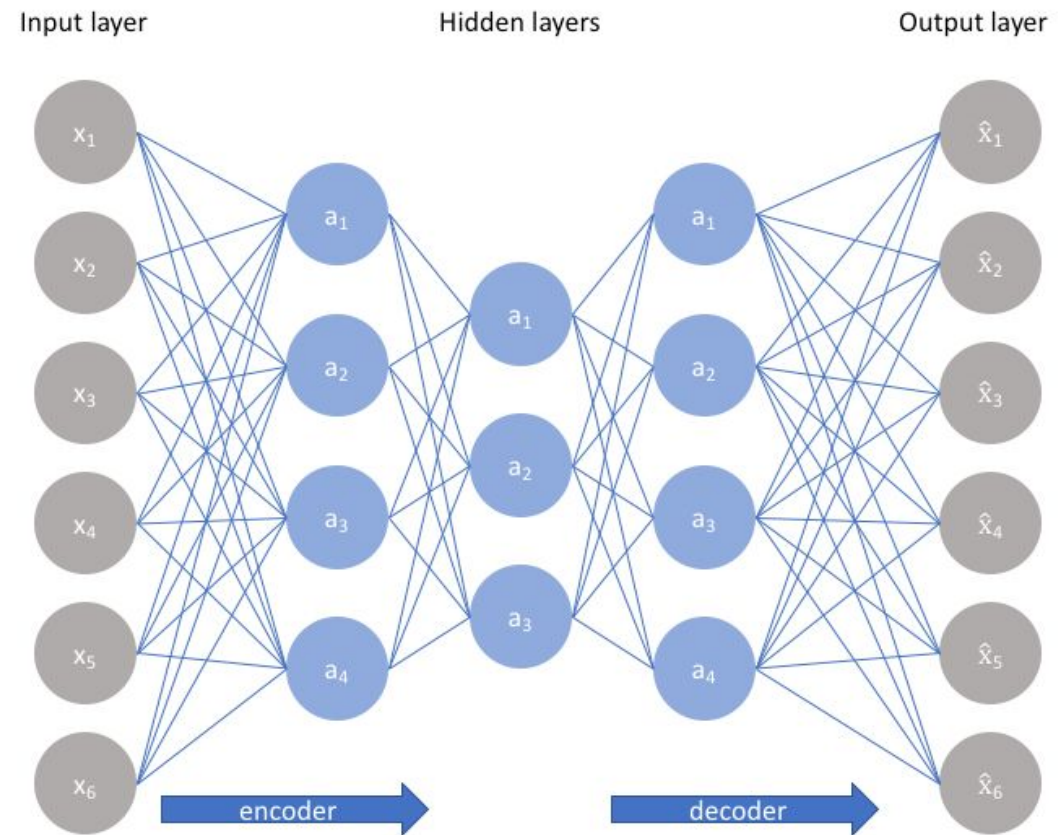


Image: CERN

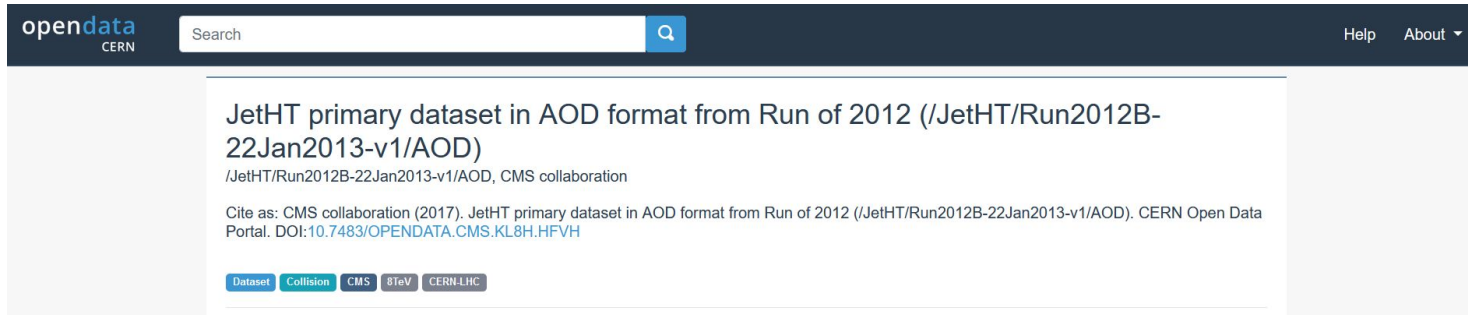
AUTOENCODERS

Autoencoders architecture used:

input-200-100-50-**15**-50-100-200-output



DATA AND EVALUATION



The screenshot shows the CERN Open Data portal interface. At the top left is the 'open data CERN' logo. A search bar is located in the top center, and 'Help' and 'About' links are on the top right. The main content area displays the title 'JetHT primary dataset in AOD format from Run of 2012 (/JetHT/Run2012B-22Jan2013-v1/AOD)' and the subtitle '/JetHT/Run2012B-22Jan2013-v1/AOD, CMS collaboration'. Below this is a citation: 'Cite as: CMS collaboration (2017). JetHT primary dataset in AOD format from Run of 2012 (/JetHT/Run2012B-22Jan2013-v1/AOD). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.KL8H.HFVH'. At the bottom of the content area are several tags: 'Dataset', 'Collision', 'CMS', '8TeV', and 'CERN-LHC'.

Compact Muon Solenoid experiment at CERN dataset

Evaluation metrics

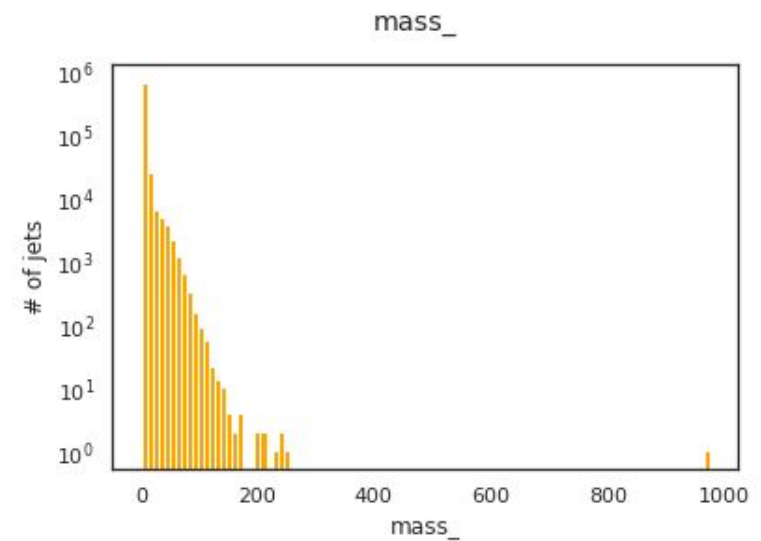
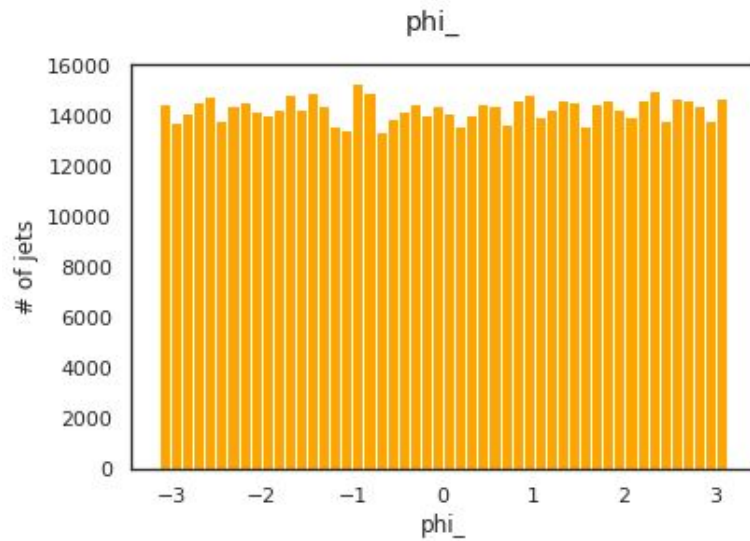
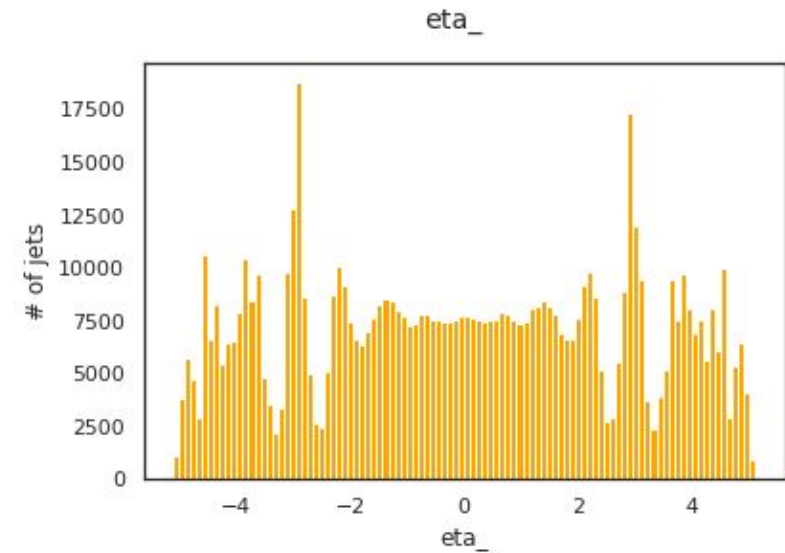
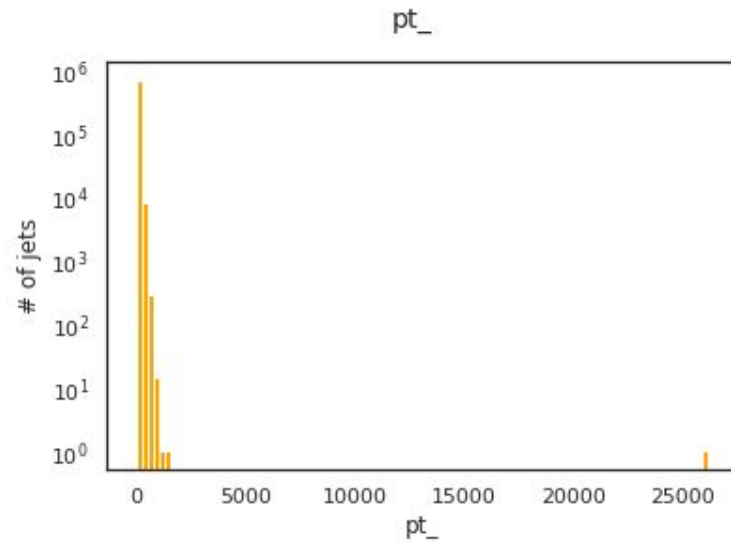
Residuals:

$$X_{in} - X_{out}$$

Relative residuals/response:

$$\frac{X_{in} - X_{out}}{X_{in}}$$

DATA AND EVALUATION



TIMELINE & FURTHER USE

Timeline

- June 20th - July 4th: project proposal.
- July 4th - July 18th: running ML algorithms on laptop and turning them into a Jupyter notebook, comparison with PCA.
- July 18th - August 20th: implementing the autoencoder on the EOSC resources.
- August 20th - August 31st: wrap up (writing report, documentation).

Further use

- Possibility to use Autoencoders for other experiments and fields
- To use Autoencoders for anomaly detection



THANK YOU FOR ATTENTION!

Student: Maxym Naumchyk