

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

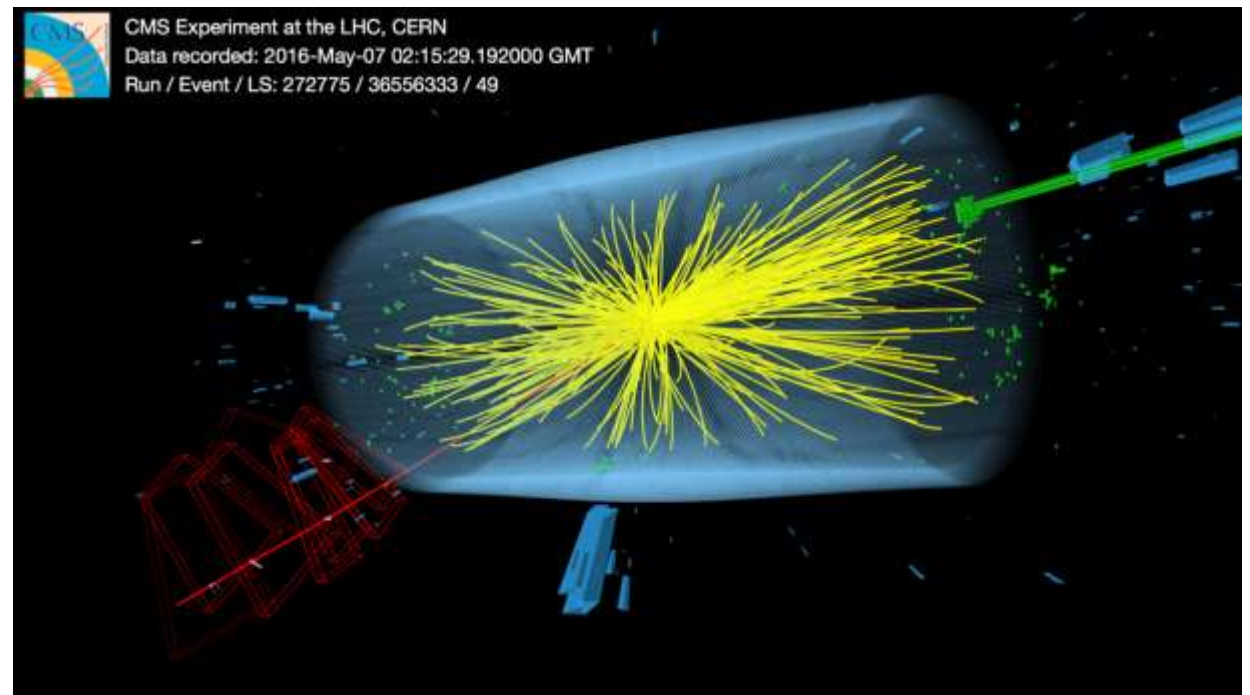
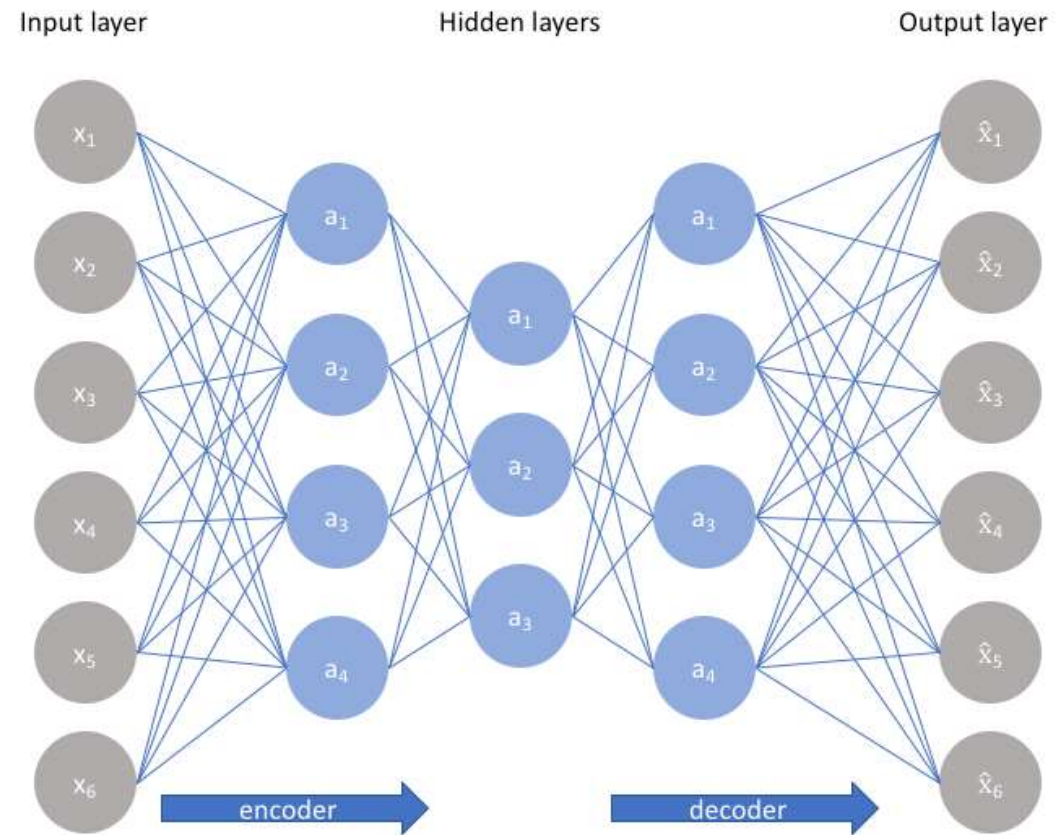


Image: CERN

AUTOENCODERS

Autoencoders architecture used:

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



DATA AND EVALUATION



Compact Muon Solenoid experiment at CERN dataset

Evaluation metrics

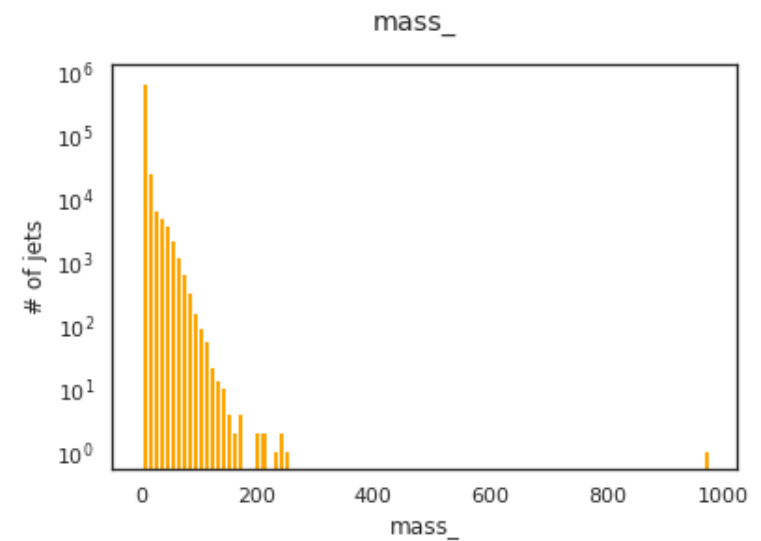
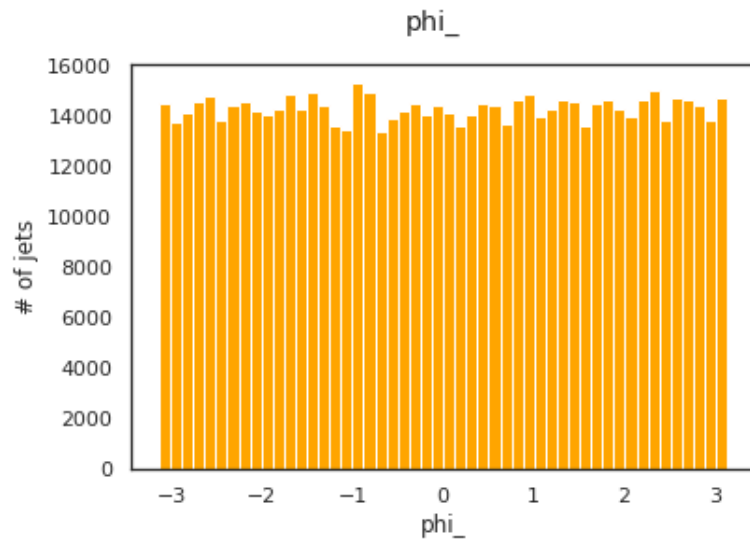
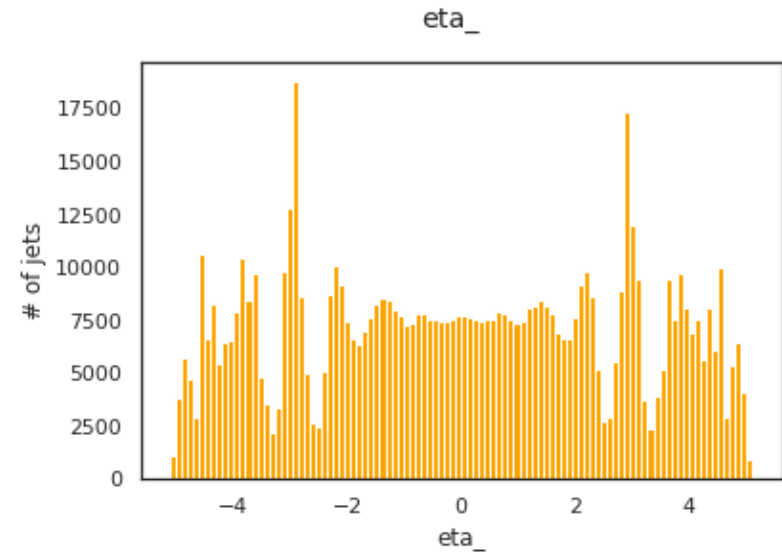
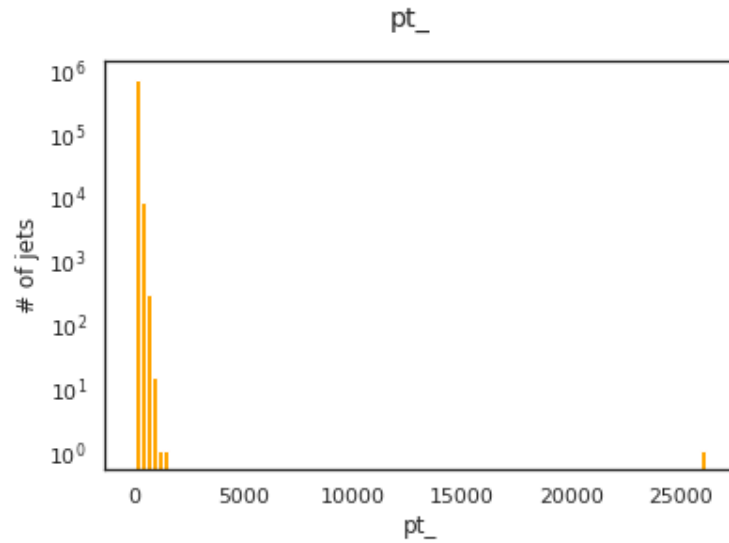
Residuals:

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

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