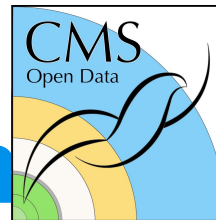
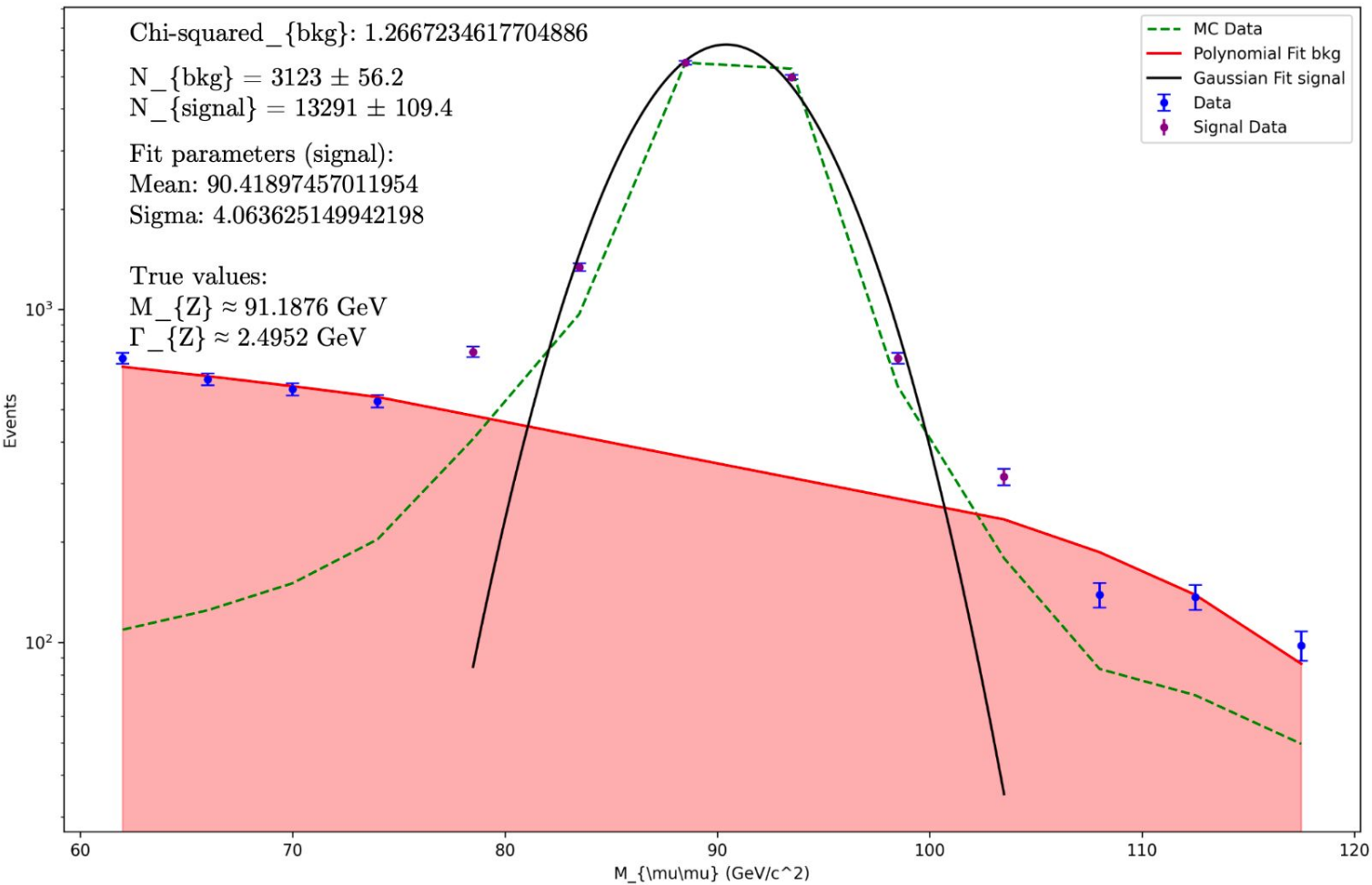
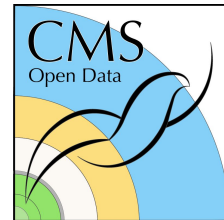


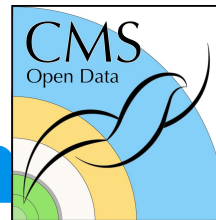
CMS Open Data Workshop 2024 Hackathon



Aug 1, 2024
Iurii Korsakov
Particle Discovery Lab



CMS Open Data Workshop 2024 Hackathon

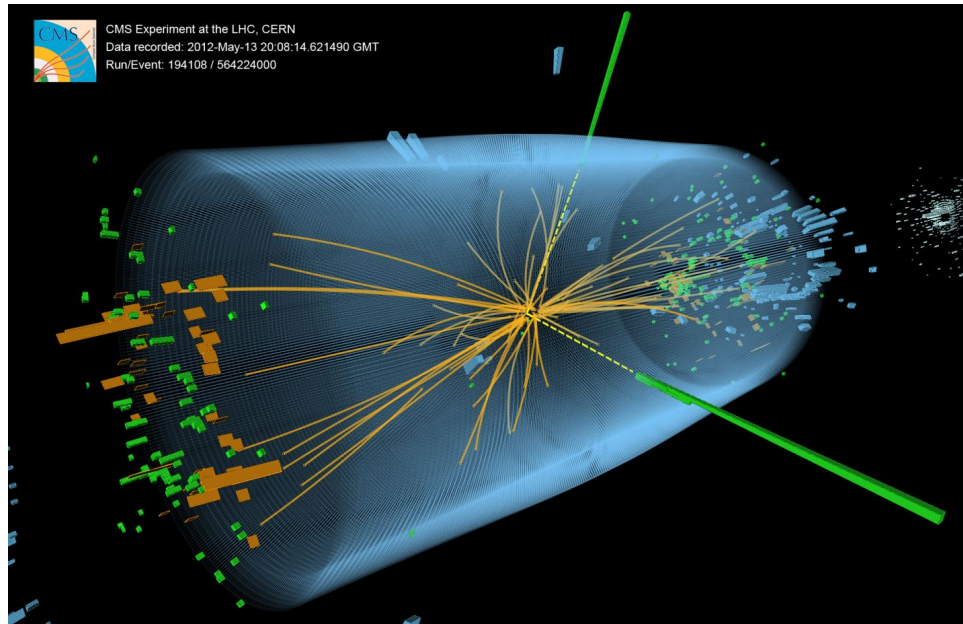


Aug 1, 2024

Camila Ramos - Piotr Sobczynski - Xavier Tintin

Machine Learning in HEP

Anomaly detection in HEP



Unsupervised Learning
Autoencoders

• Autoencoders (AE)

Steps:

1. **Data Preparation:**
 - a. **Fetch Data**
 - b. **Read Data**
 - c. **Convert to Regular Arrays**
 - d. **Standardize Data**
2. **Autoencoder Definition:**
 - a. **Architecture**
 - b. **Training**
3. **Evaluation:**
 - a. **Reconstruction**
 - b. **Anomaly Detection**
 - c. **ROC Curve**

• Autoencoders (AE)

Data Preparation

Data input

- **Background:**
 - 2,14 M events
 - MET, Electrons, Jets, and Muons
 - p_T , η , φ , and particle ID

- **Signal**
 - 135K events
 - $\phi \rightarrow \mu\mu t$

Train

Evaluate

• Autoencoders (AE)

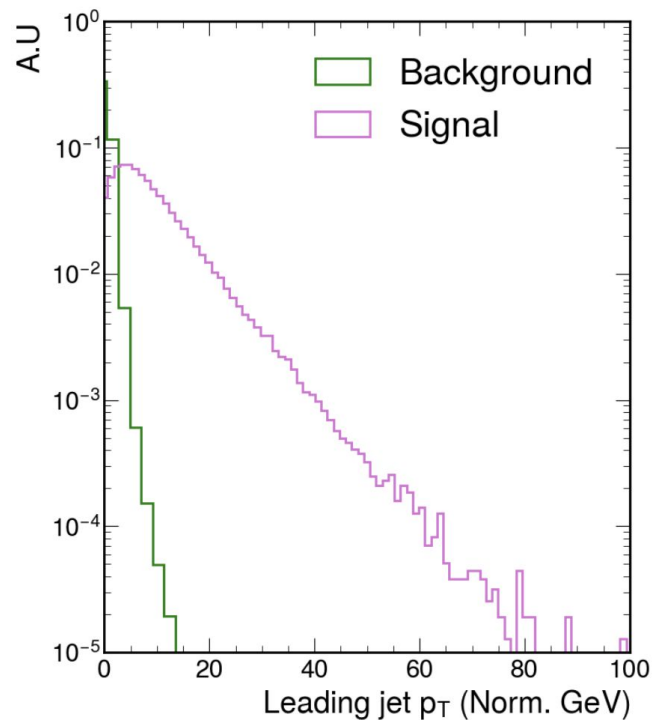
Data Preparation

Entries:

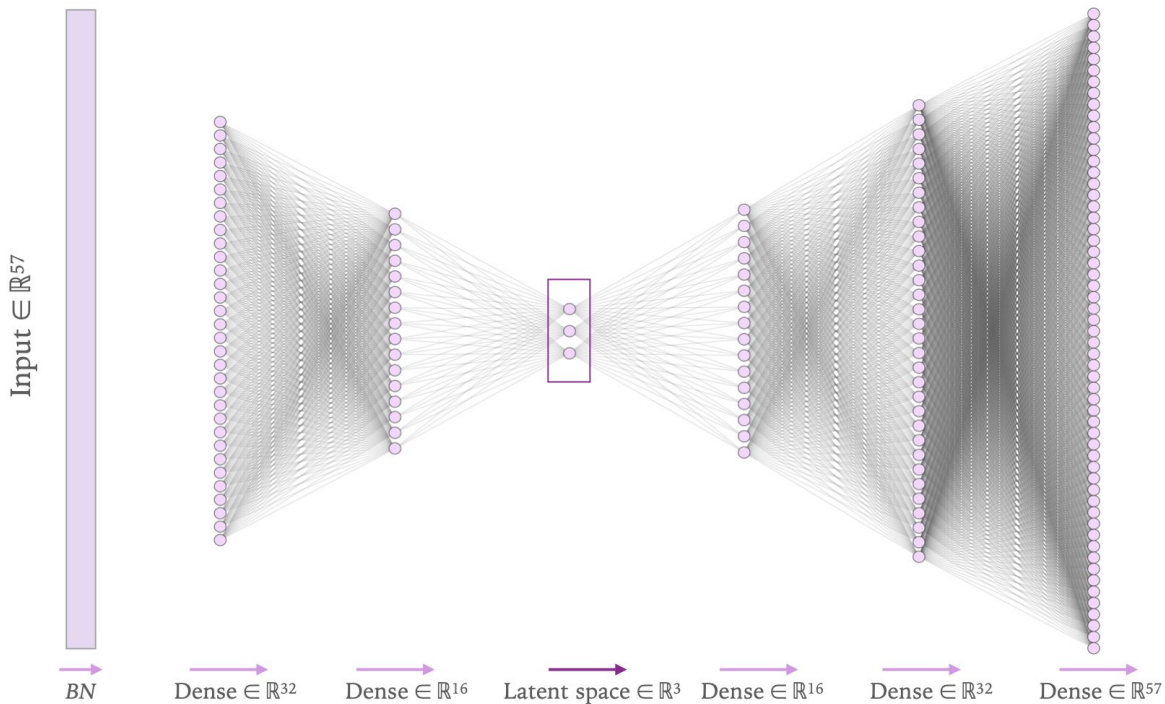
- Electrons
- Jets
- Muons
- MET

Used features:

- p_T
- η
- ϕ



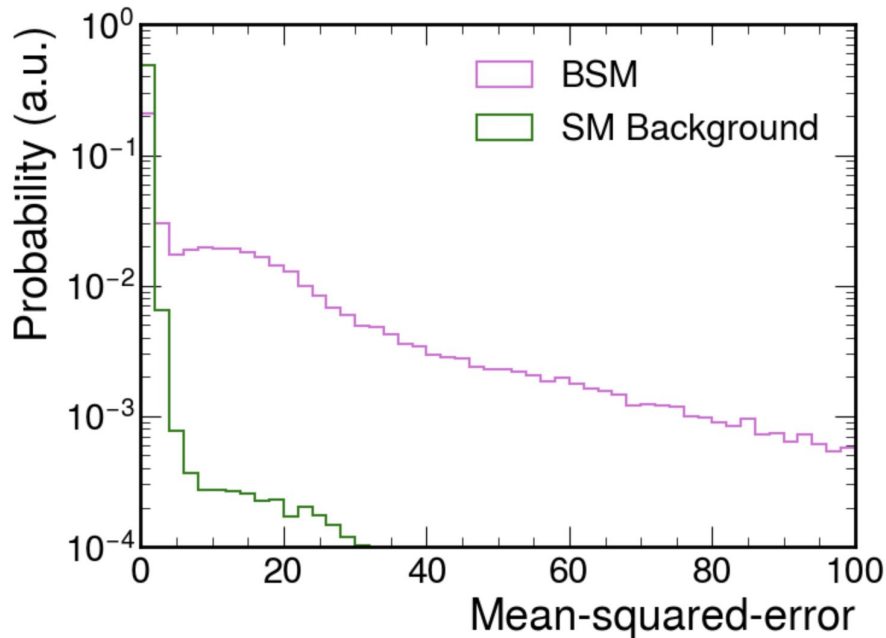
- **Autoencoders (AE)**
Autoencoder Definition



• Autoencoders (AE) Evaluation

Key Metric:

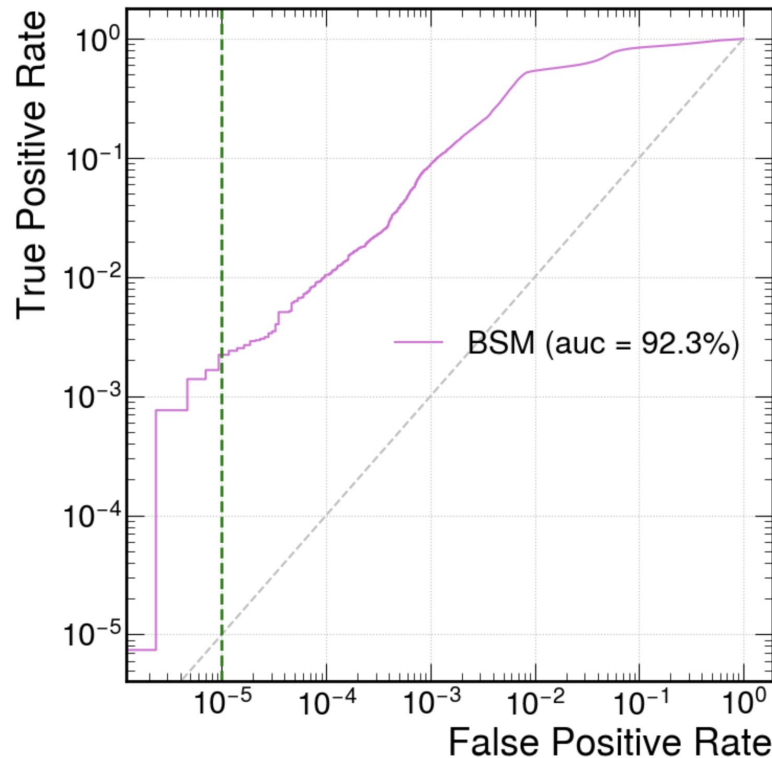
Anomaly detection is the mean-squared-error: If the error is high, the data is more likely to be anomalous.



• Autoencoders (AE) Evaluation

ROC:

- The TPR is the fraction of anomalous events that are correctly identified by the model.
- The FPR is the fraction of normal events that are incorrectly identified by the model as anomalous.
- The vertical line at $FPR = 0.00001$ shows a possible threshold for anomaly detection. Events with a reconstruction error (mean squared error) above this threshold would be classified as anomalous.





Thank you!