Measuring the Relative Performance of Linac Beam Outages with L-CAPE Anomaly Detection

Brillina Wang, Elgin Community College – Jason St. John, Brian Schupbach

FERMILAB-POSTER-22-198-STUDENT

Introduction

- Linear Accelerators (Linacs) accelerate charged subatomic particles and subjects them to a series of electric potentials along a linear beamline. They are mainly composed of radiofrequency (RF) cavities placed in-line with one another to provide a large amount of energy gain per unit length.
- Machine learning (ML) refers to the training of machines to get better at a task without explicit programming. Anomaly detection is a ML method of identifying unexpected events or observations that differ significantly from the predicted value (see Figure 1). At the Fermilab AD division, L-CAPE is focused on predicting and preventing Linac beam outages with anomaly detection (see Figure 2).

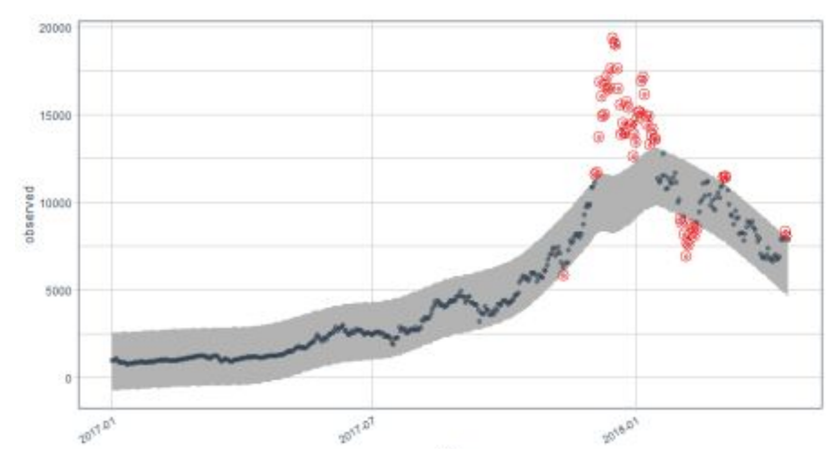

Figure 1: Anomaly Detection Plot

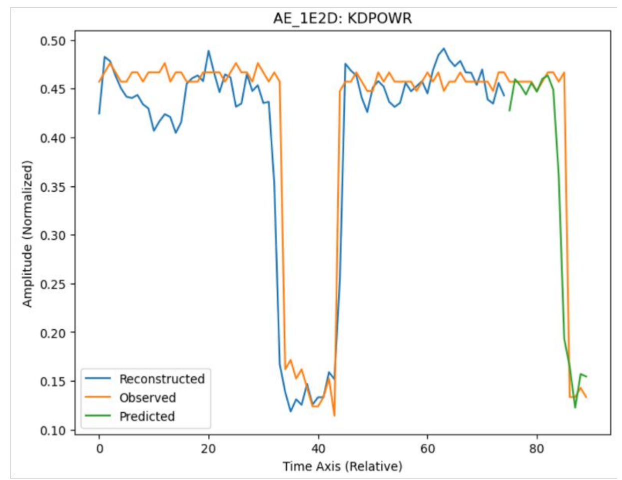

Figure 2: PNNL Private Correspondence Autoencoder

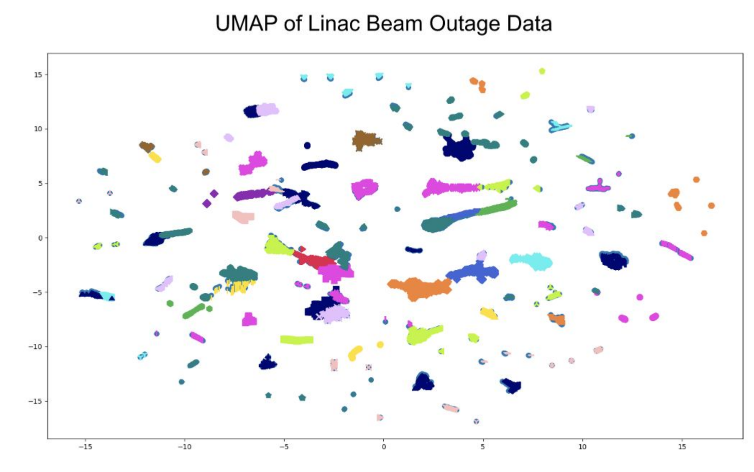

Figure 3: UMAP to visualize clustering patterns in high-dimensional data, data points colored by ground truth fault type labels

Methods

- The readings of the Linac control system devices constitute a time series of numbers. The data was prepared for UMAP by taking ten-second intervals for 24 devices, making a 10x24 dimensional space in 2D (see Figure 3). Clustering will not work in high dimensions due to the curse of dimensionality, which refers to the phenomenon that the number of samples required to estimate an arbitrary function with a given level of accuracy grows exponentially with respect to the number of input variables.
- Two clustering algorithms were implemented using the scikit-learn built-in machine learning library for the Python programming language.
  - **Gaussian Mixture Method (GMM)**: A probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters (see Figure 4).
  - **Hierarchical Agglomerative Clustering (HAC)**: A method of cluster analysis that seeks to build a hierarchy of clusters (see Figure 5).
- Within the same library, effectiveness of each algorithm was primarily based on two metrics.
  - **Homogeneity score**: Measures the purity of clusters in terms of fault type.
  - **Completeness score**: Measures how complete each fault type is contained by single clusters.
- Two sets of scores were obtained for HAC algorithm, one with no clusters set and one that asks to find 59 matched clusters, due to there being 59 unique fault types.

Purpose

To study the relative performance of these clustering algorithms on high-dimensional data downprojected to low-dimensional space (Uniform Manifold Approximation and Projection, UMAP).

Results & Conclusions

- **GMM**
  - Homogeneity score: 0.772
  - Completeness score: 0.619
- **HAC (optimized clusters, None)**
  - Homogeneity score: 0.999
  - Completeness score: 0.326

HAC (59 matched clusters)
- Homogeneity score: 0.783
- Completeness score: 0.633

For these scores, a 1 indicates a perfect score and a 0 indicates an imperfect score. When working with 59 clusters, the homogeneity and completeness scores of HAC are closer to 1, which suggests that this algorithm is more effective in grouping unusual data points. In the future, we are hoping to be able to calculate the fractional probabilities of the possible fault types, the percent error of predictions, as well as the reason why some fault types overlap with others on the UMAP.

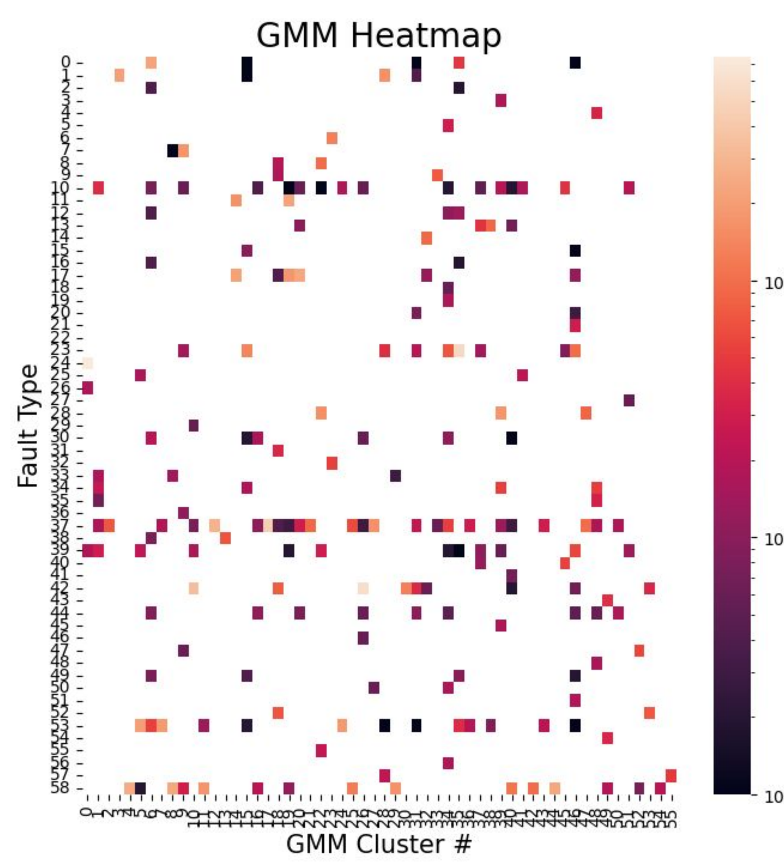

Figure 4: Heatmap Representation w/ GaussianMixture Method of Fault Type & Cluster #

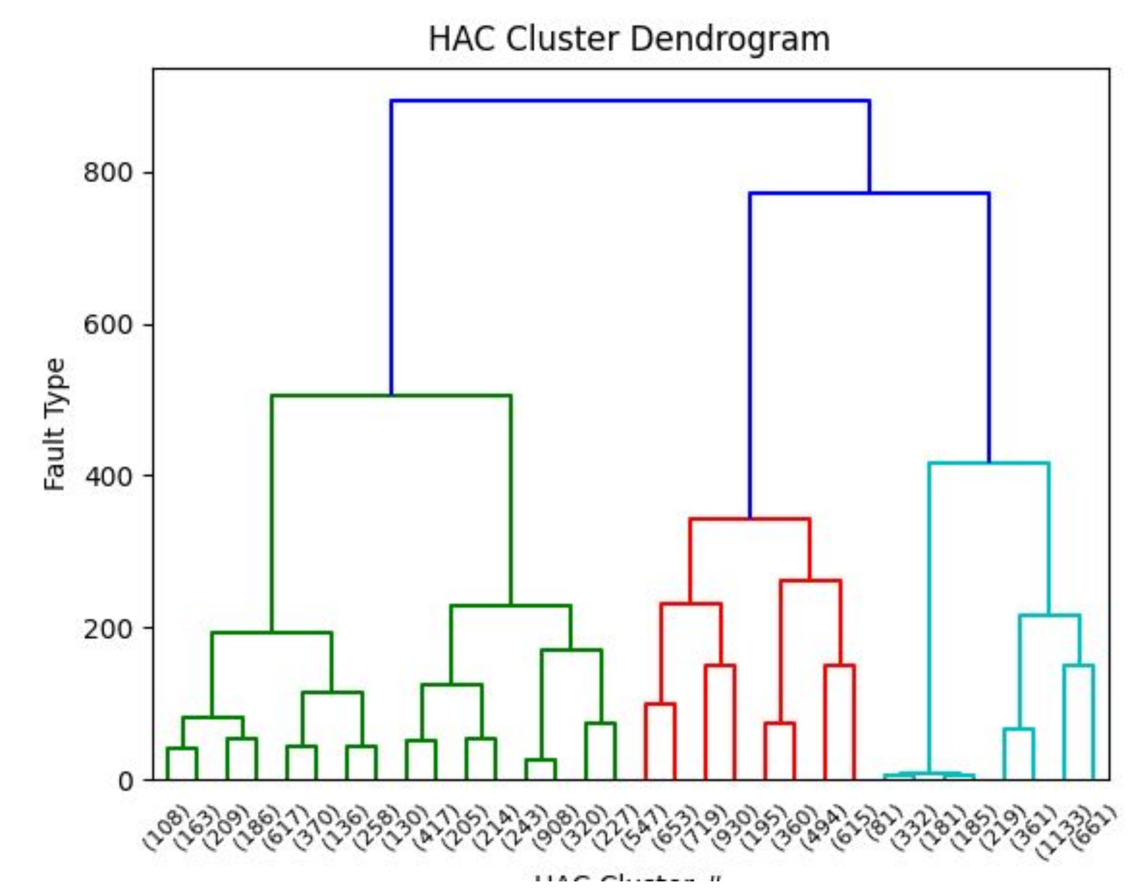

Figure 5: Dendrogram Representation w/ AgglomerativeClustering Method of Fault Type & Cluster #

Acknowledgements

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.