



# Enhancing Electronic Logbooks Using Machine Learning

Jennefer Maldonado, Samuel Clark  
Collider Accelerator Department, Brookhaven National Laboratory

[jmaldonad@bnl.gov](mailto:jmaldonad@bnl.gov)

[sclark@bnl.gov](mailto:sclark@bnl.gov)

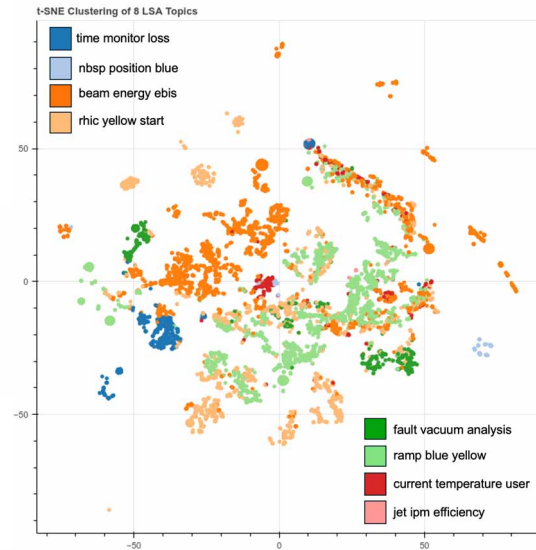


# Overview

- NLP techniques are applied to all entries in the elog DB
  - Processing data, removing stop words, lemmatizing
- Doc2Vec and Multinomial Naïve Bayes
- Feedback suggested researching two models Latent Semantic Analysis and Latent Dirichlet Allocation
- Web Based Search Engine
  - Connected to the model class

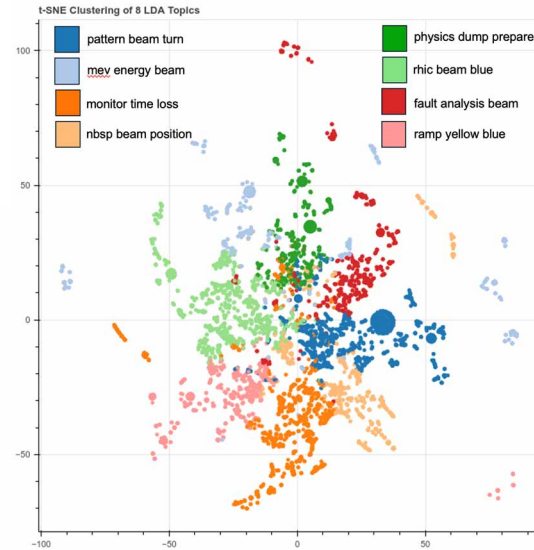
# Latent Semantic Analysis

- Truncated Singular Value Decomposition of the document-term matrix
- Linear dimensionality reduction



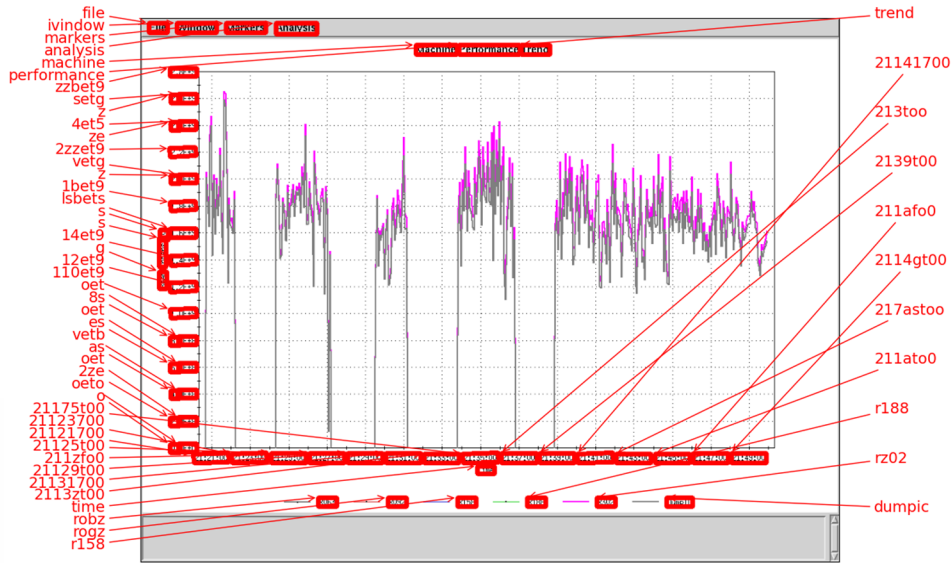
# Latent Dirichlet Allocation

- Probabilistic process
- Ratio count vectors



# Optical Character Recognition

## Keras-OCR



File Window Markers Analysis

Machine Performance Trend

