This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under GA No 101004730.

Data analysis for the accelerator domain

I.FAST Task 10.6

Karlis Berkolds (RTU)

Irena Dolenc Kittelmann (ESS), Agris Nikitenko (RTU), Thomas Shea, (ESS)

iFAST

3rd I.FAST Annual Meeting 2024-04-18

10.6 Task Summary

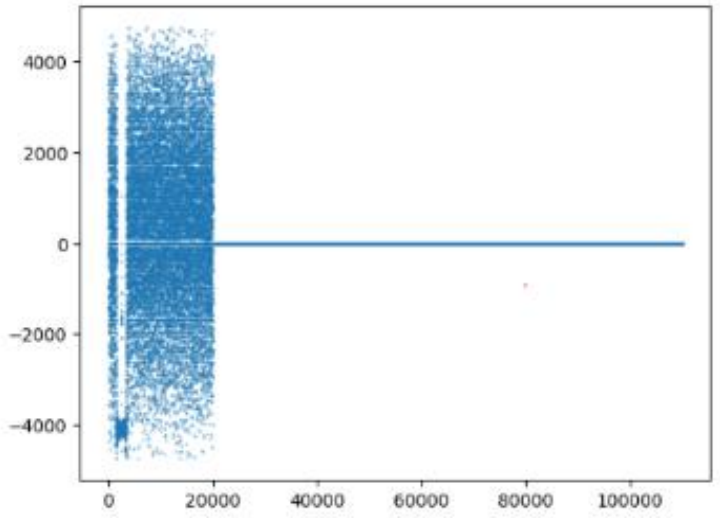
- Long term mission:
  - Develop low-latency Machine Learning (ML) techniques to improve performance and availability of high-power facilities at the intensity frontier
- Goal:
  - Identify signatures of potential errant beam conditions
- Scope:
  - Identify relevant data
  - Assess the predictive capabilities of selected ML models
  - Prototype: proof of principle demonstration
    - The most promising ML model to be implemented on a low-latency network of FPGAs processing signals from an array of detector channels

Data and prototyping setup

- Data acquired during commissioning of the ESS normal-conducting linac:
  - ESS Beam commissioning through DTL 4 (74 MeV)
  - Scalar data and waveform data from:
    - Beam Current Monitor (BCM)
    - Beam Position Monitor (BPM)
- Software: Python, Jupyter Notebooks, VS code
- Hardware: Local machine
- Challenges:
  - Limited Archiver capabilities
  - Data downloading speed
  - Data downloading format
  - Data query size
  - Local storage and retrieval

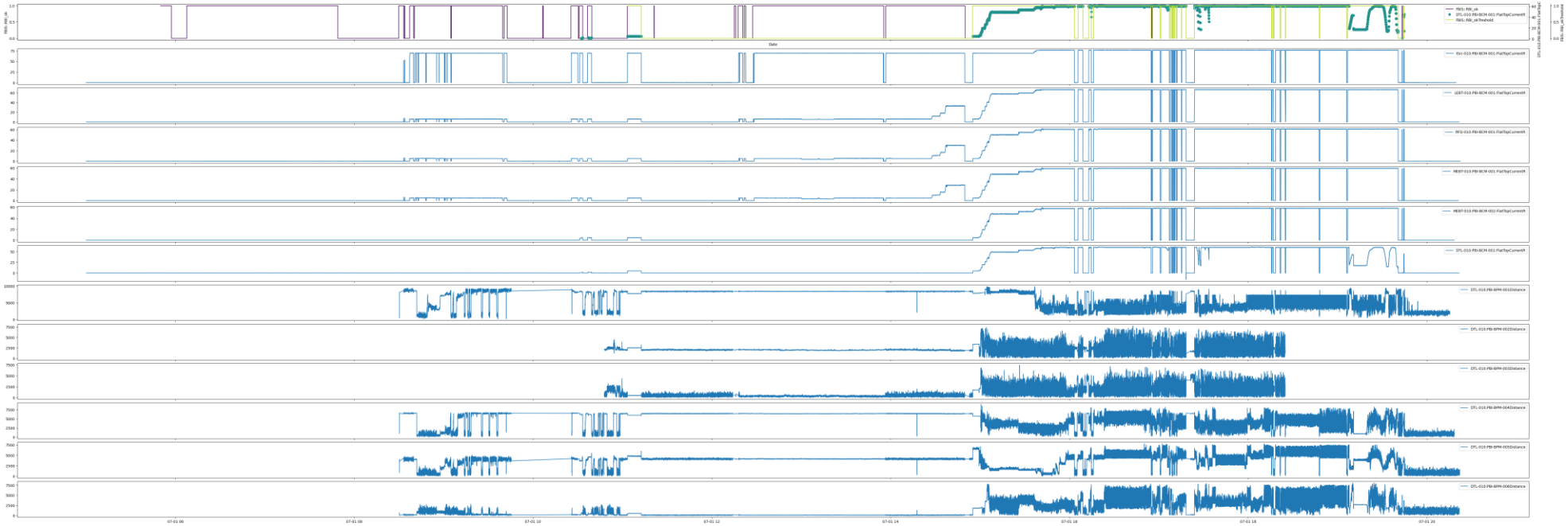
Processing Waveforms

- Usable count: ~12 000 machine cycles with beam pulse present
- Full Waveform size: 110 000 points
- Cut to leave only the representative part
  - ~3000 points



Filtering cycles with beam present

- Before the interlock
- Representative beam current
- Beam current value threshold

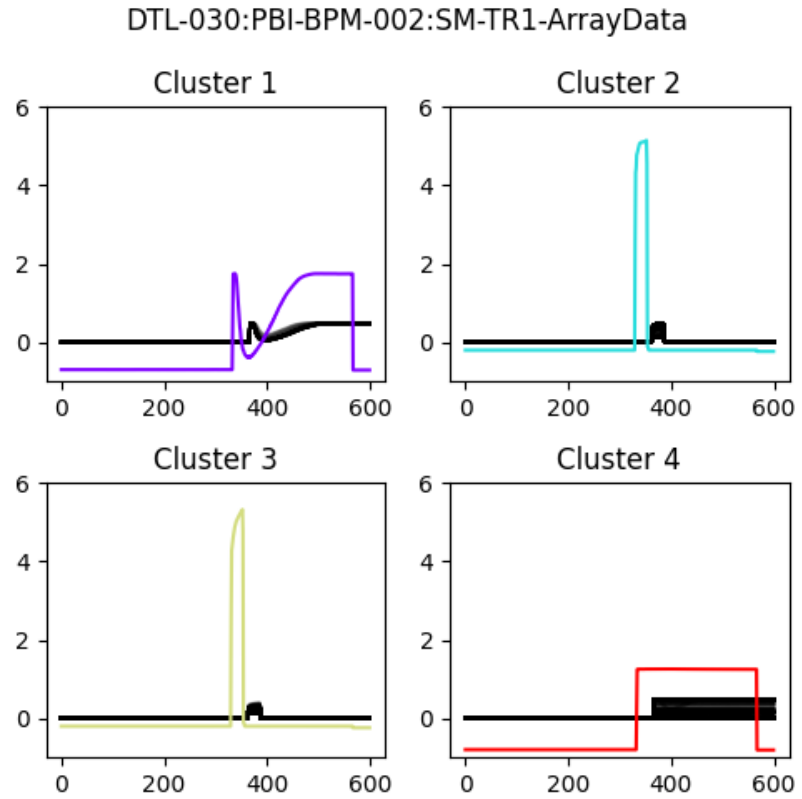


Clustering

- Grouping unlabeled examples
  - Relies on unsupervised machine learning
- Example feature data combined into similarity measure.
  - Similar examples grouped
- Approaches
  - Centroid-based clustering
  - Density-based clustering
  - Distribution-based Clustering
  - Hierarchical Clustering
  - Others…

Kshape Clustering

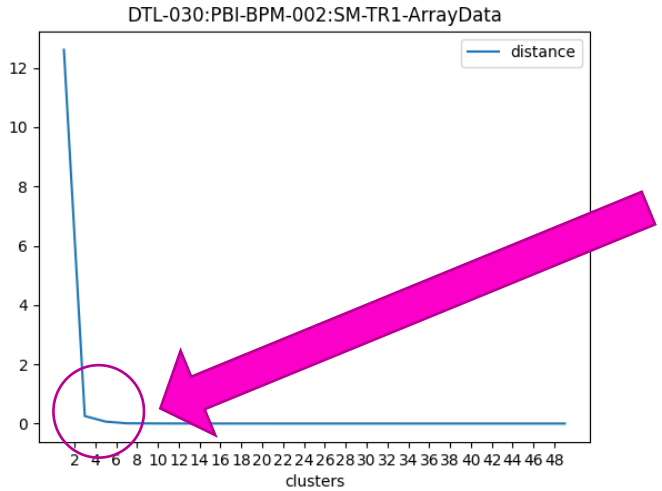
- Selection:
  - Waveforms are time series
  - Waveforms can be shifted in time
- K-Shape clustering
  - Clusters Waveforms according to shape
  - Using Dynamic Time Warping
    - Similarity measure includes time shift
  - Requires: cluster count
  - Outputs:
    - Cluster index for each example
    - Cluster center shape



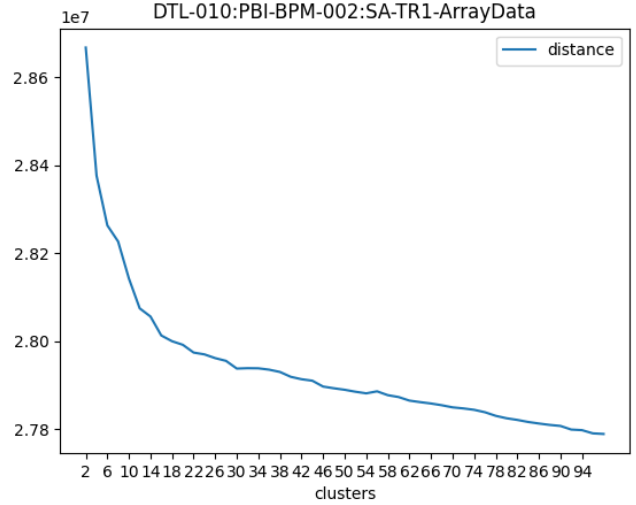
Estimating cluster count -Elbow method

- Graphically find the optimal cluster count- k value.
- Each cluster count calculates the Sum of the square distance between points in a cluster and the cluster centroid.
- Pick the value from the “Elbow point.”

BPM Phase data:
There is no clear Elbow point

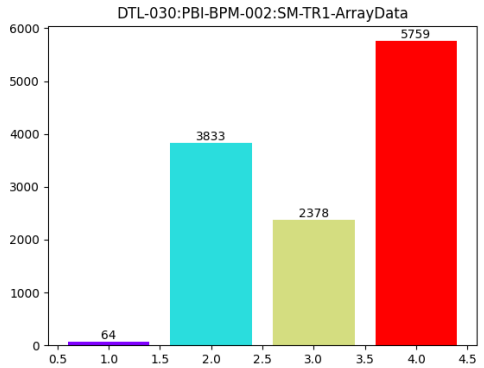


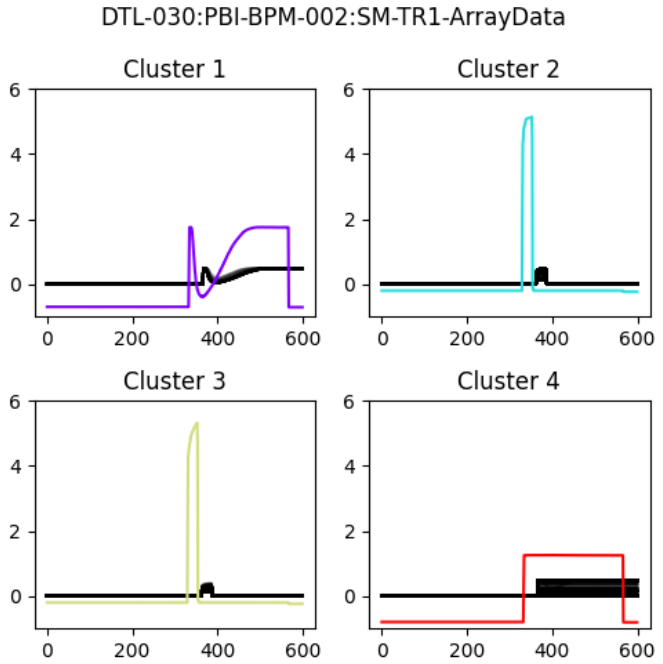
BPM Magnitude data;
Clear Elbow point

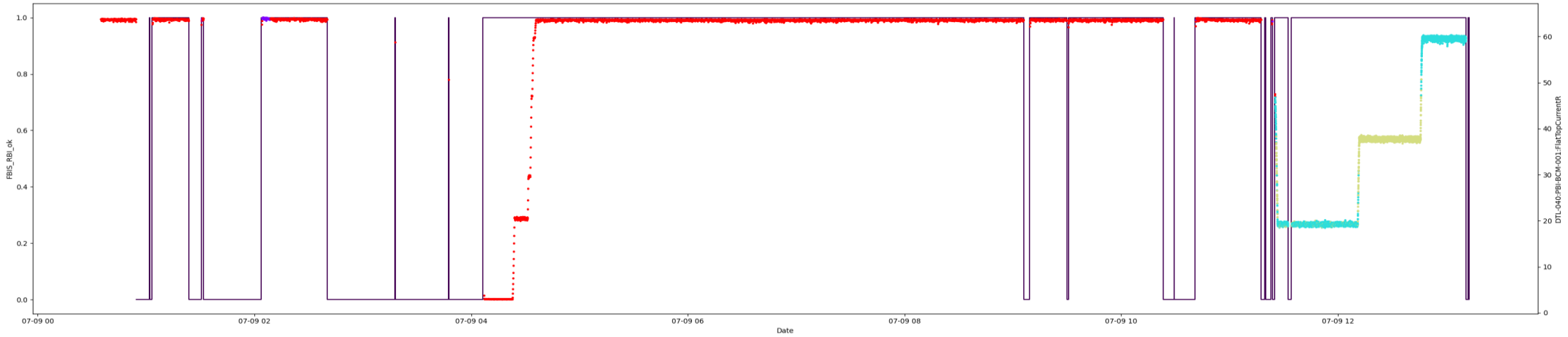


Searching for events before interlocks

- Clustering each individual waveform
- Determine cluster centre waveform
- Combine:
  - Clustering result
  - Current value
  - Interlocks

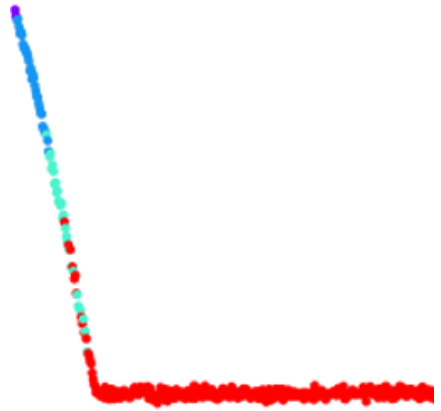




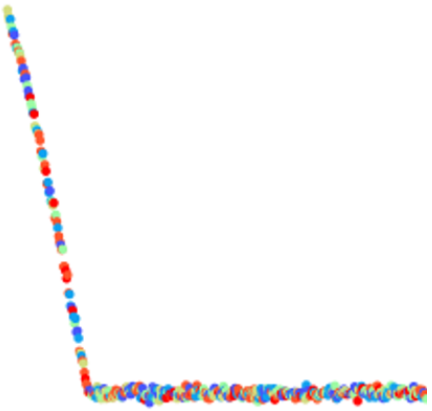


Initial results

- BPM Magnitude data sets:
  - Clusters waveforms
  - Some correlations with the current value
  - No apparent waveform change before the interlock



- BPM Phase data sets
  - No distinct waveforms with the k-shape method



Correlations between systems and/or data variables

Current work – Random forest method

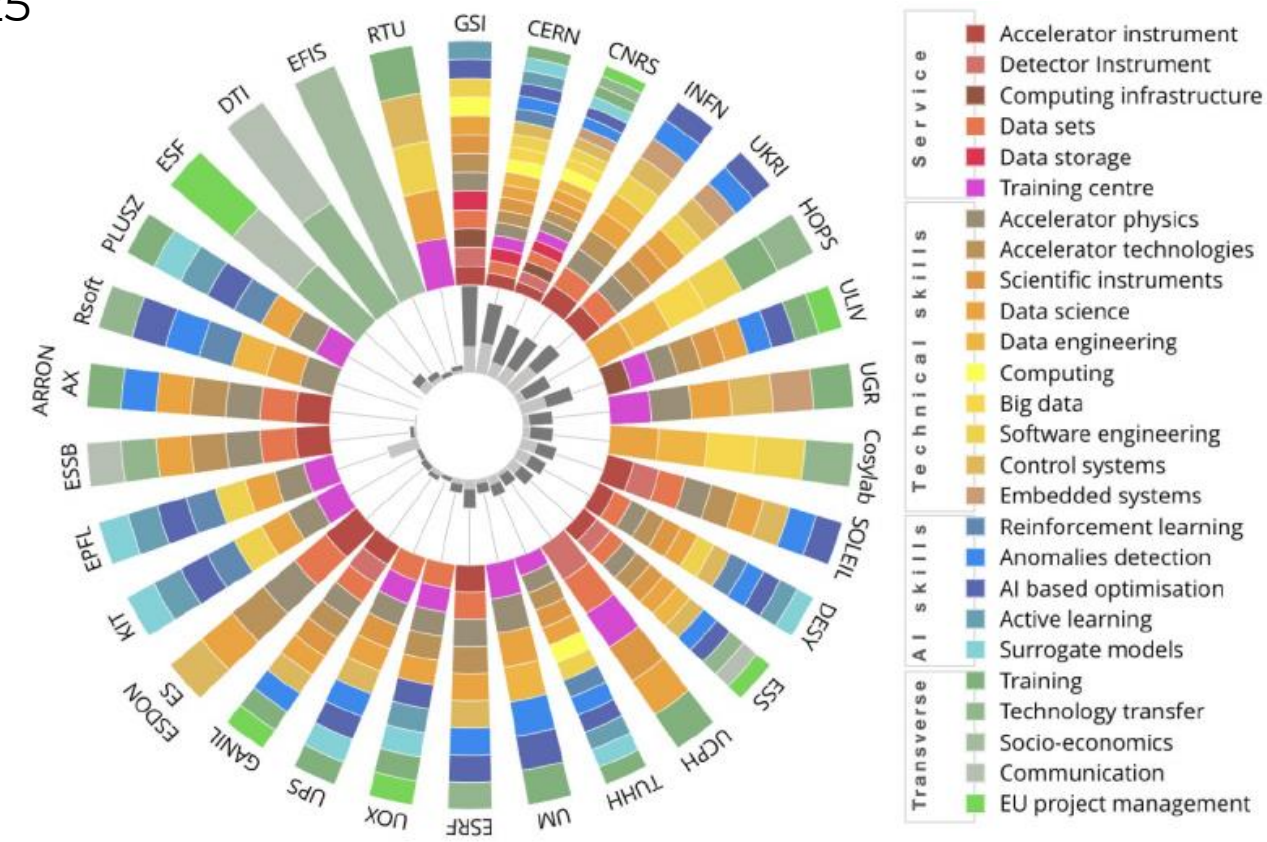
- Select a random machine cycle set with waveforms from different systems
- Use a random subset of waveforms to build decision trees

- Random Forest is created
  - Pass the data through the random forest to get predictions on the next waveform value.
  - Aggregate individual predictions to get the new value
- If there is a difference between prediction and actual value, there might be an **Event**

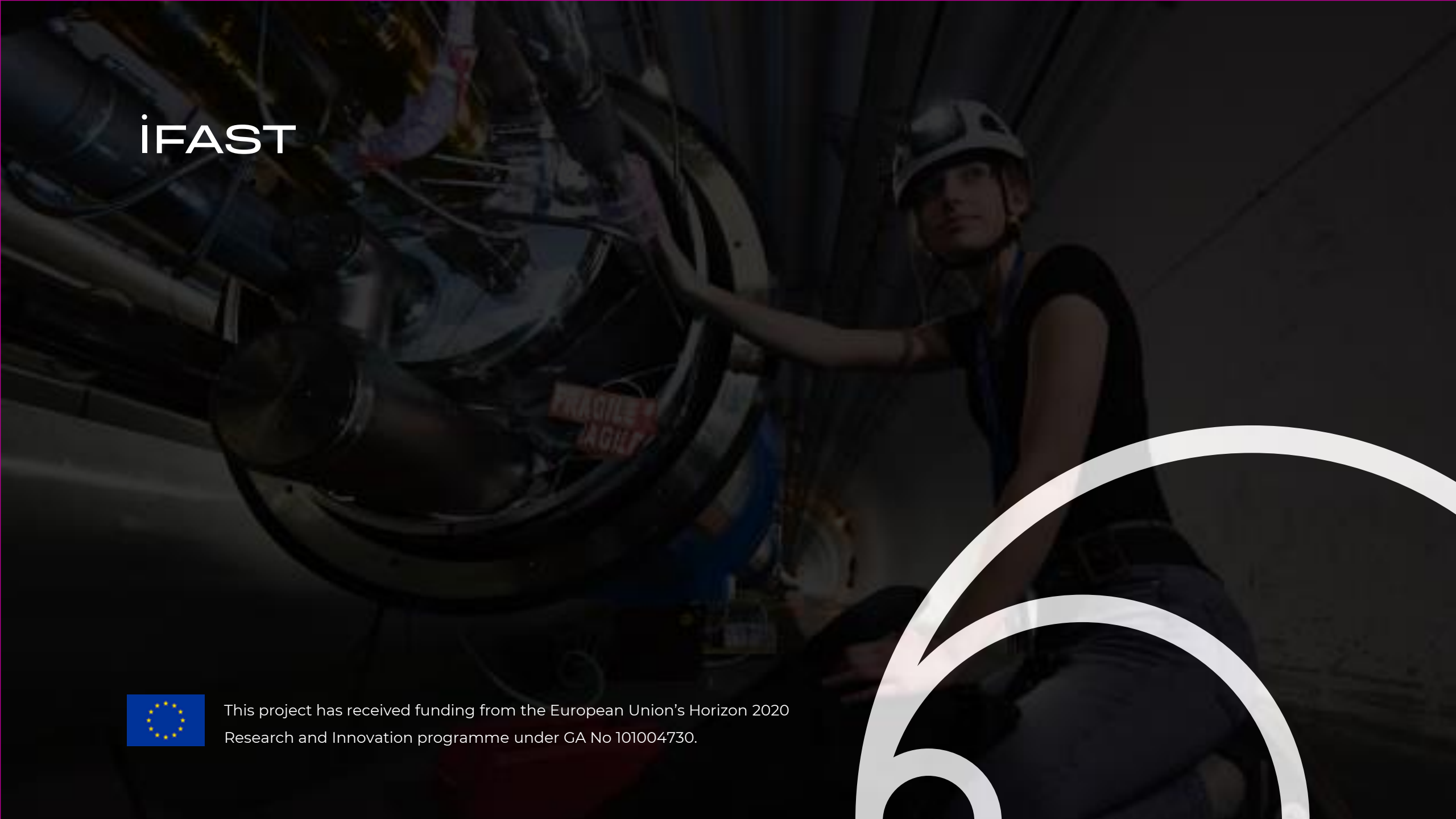
ARTIFACT

ARTificial Intelligence For Accelerators, user Communities and associated Technologies

- Horizon Europe proposal submitted in March
  - 4 year program startup planned 2025
  - 10MEUR
- Over 30 EU labs and organisations, GANIL, CERN, ESS, DESY, RTU, …
- 7 work packages
  - WP6: AI Algorithms and Methods
  - Task on anomalies detection identified (Task 6.3)
  - Significant experience in data formatting, handling & standardization
- Independent of funding success, a consortium in being established and collaboration is commencing



Yngve Levinsen



This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under GA No 101004730.