GEIST Research Group
We are GEIST. We dream big and work hard.

JAGIELLONIAN UNIVERSITY
IN KRAKÓW

AIRA SEMINAR

Natalia Wojak-Strzelecka
24.04.2025

Enhancing concept drift detection, explanation and adaptation to changes in industrial data streams

Table of Contents

MOTIVATION
Daily life challenges in industry

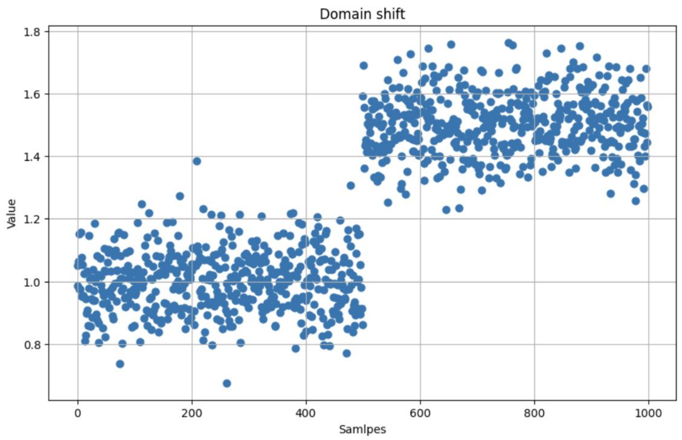
NEW
PRODUCTION
LINES

NEW/ RARE
PRODUCTS

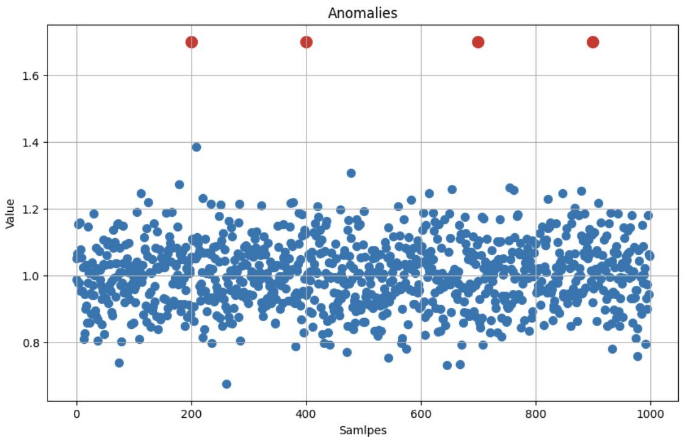
RENOVATIONS/
CALIBRATION

MOTIVATION
Challenges in industrial data streams

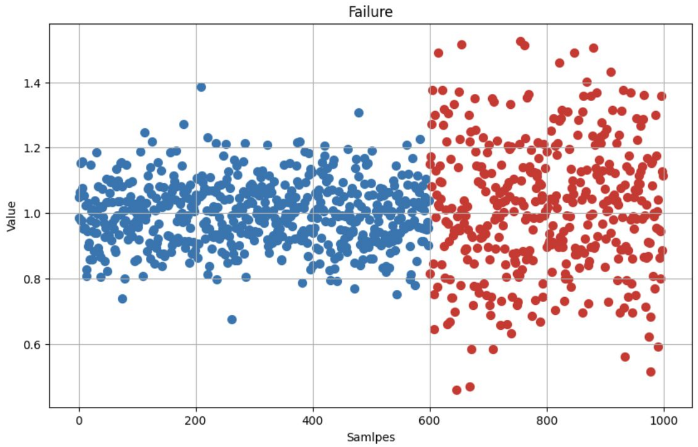
Domain Shifts



Anomalies



Failures



Problems?

1. **Domain shifts violate the i.i.d** (independence and identity) assumptions -> decrease in model performance

2. **Small domains are treated as anomalies by anomaly detectors** -> decrease in production quality, many false alarms

3. **Failures can look like domain shifts** -> no failure detection, adaptation to failures

MOTIVATION

Rapid detection and adaptation to changes in data distribution

Good representation of even small groups of data

Explanation of the changes in data streams

GEIST
Research Group

Addressing **Representation and Aggregation Biases** in Streaming Anomaly Detection through Domain Adaptation and Data Shift Detection

Natalia Wojak-Strzelecka, Szymon Bobek, Sepideh Pashami, Grzegorz J. Nalepa

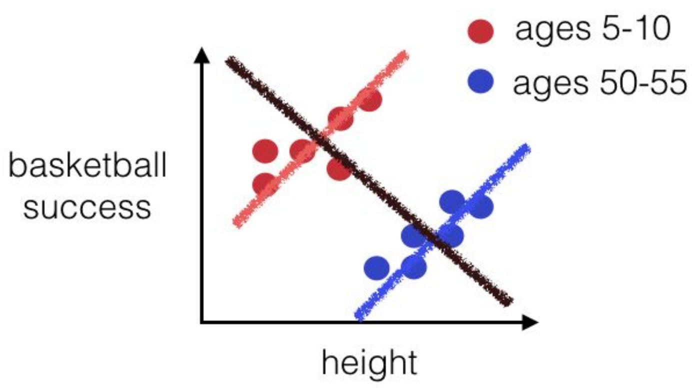
submitted to ECML2025 journal track

DEALING WITH DOMAIN SHIFTS

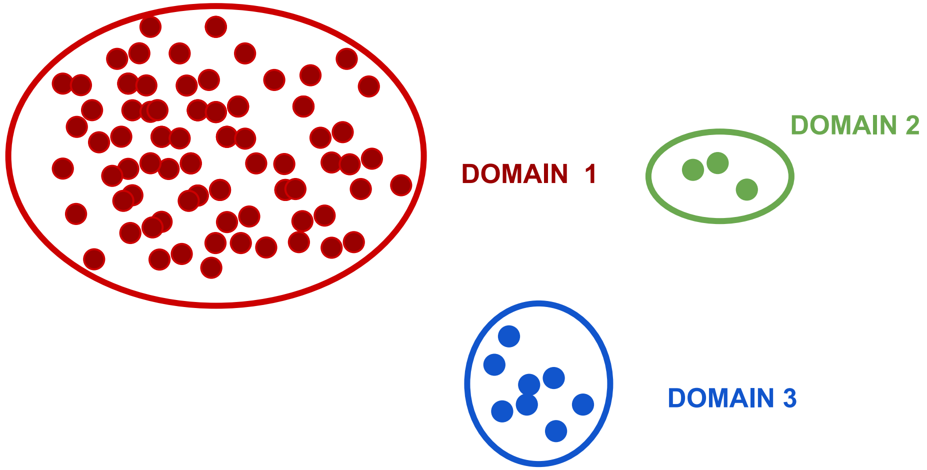
ONE MODEL FOR ALL DOMAINS

SEPARATE MODELS FOR ALL DOMAINS

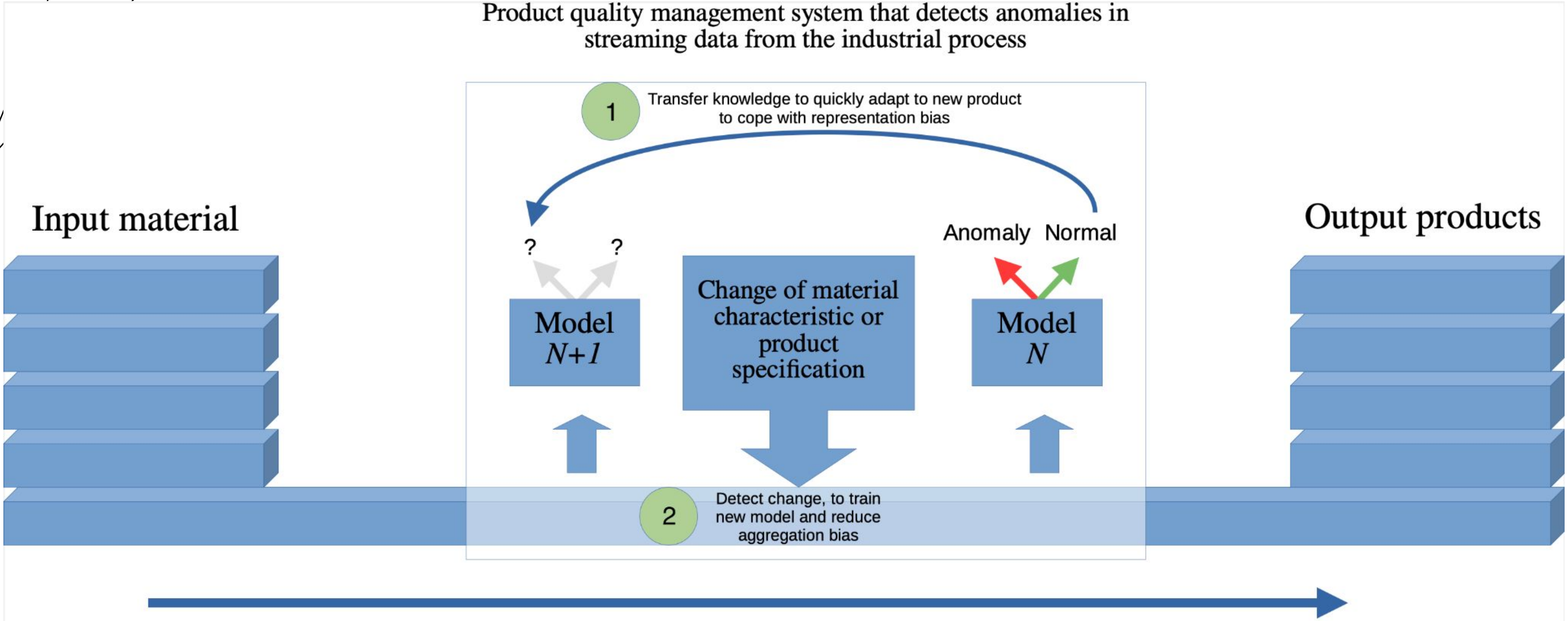
AGGREGATION BIAS



REPRESENTATION BIAS



PROPOSED APPROACH



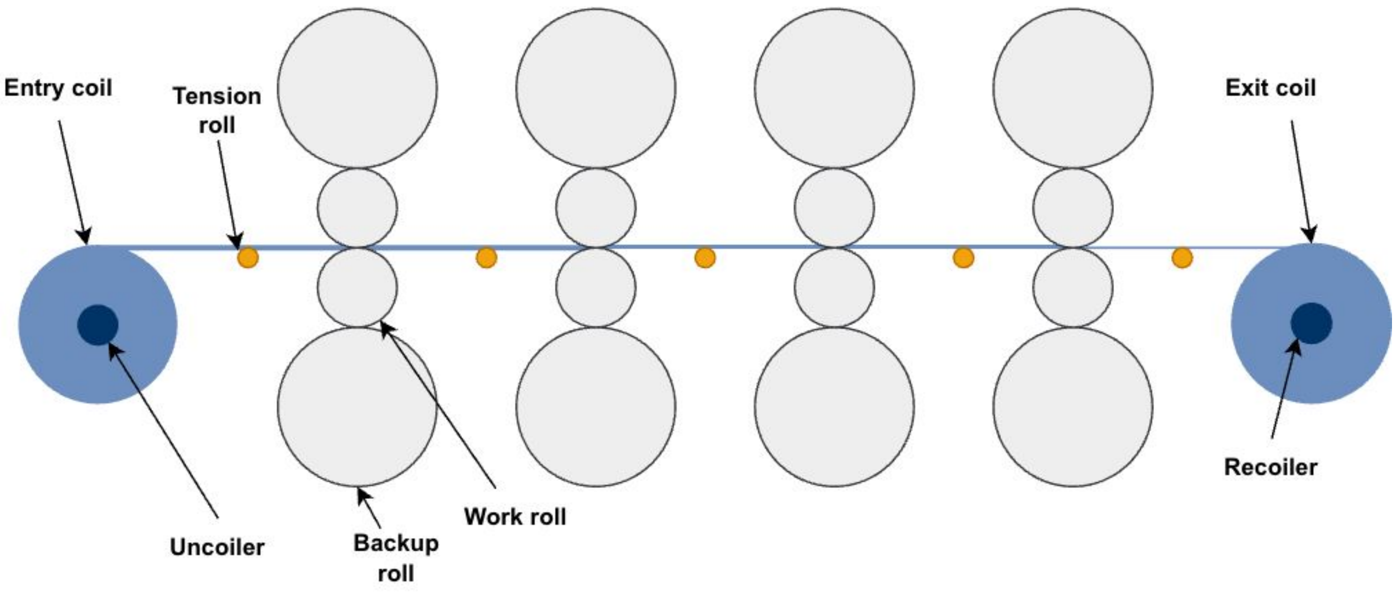
Fast production process of different types of products from differnt types of materials

1. How to deal with representation bias, where separate models for different products cannot be trained effectively due to lack of enough data
2. How to deal with aggregation bias, where one single model cannot be used to handle all types of products efficiently

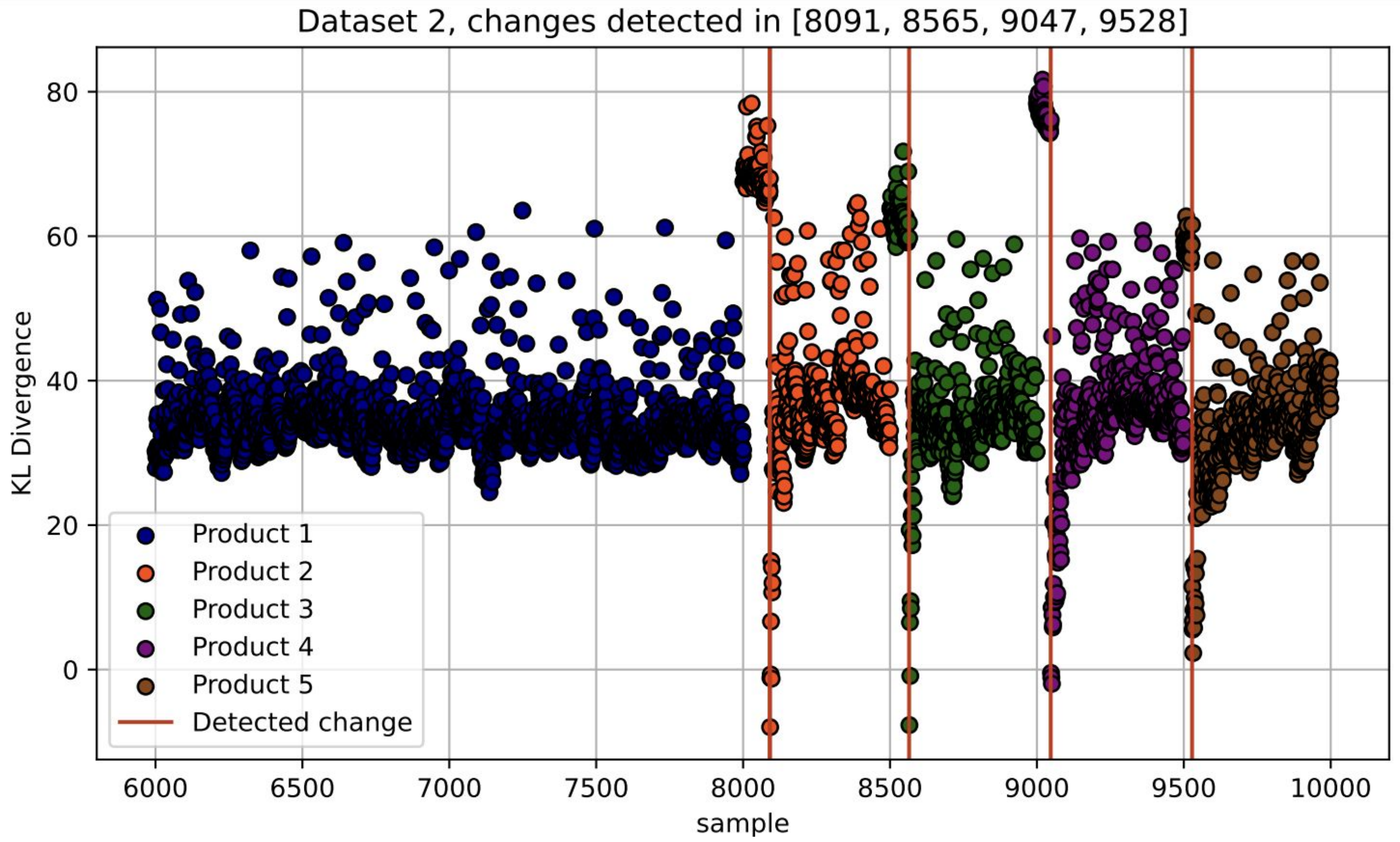
COLD ROLLING

The cold rolling process aims to reduce the steel thickness to a satisfactory level, making the product ready to be sold to customers.

- **Domain shifts** – represents a change in the data characteristics
- **Anomalies** – Short-lasting deviations in process

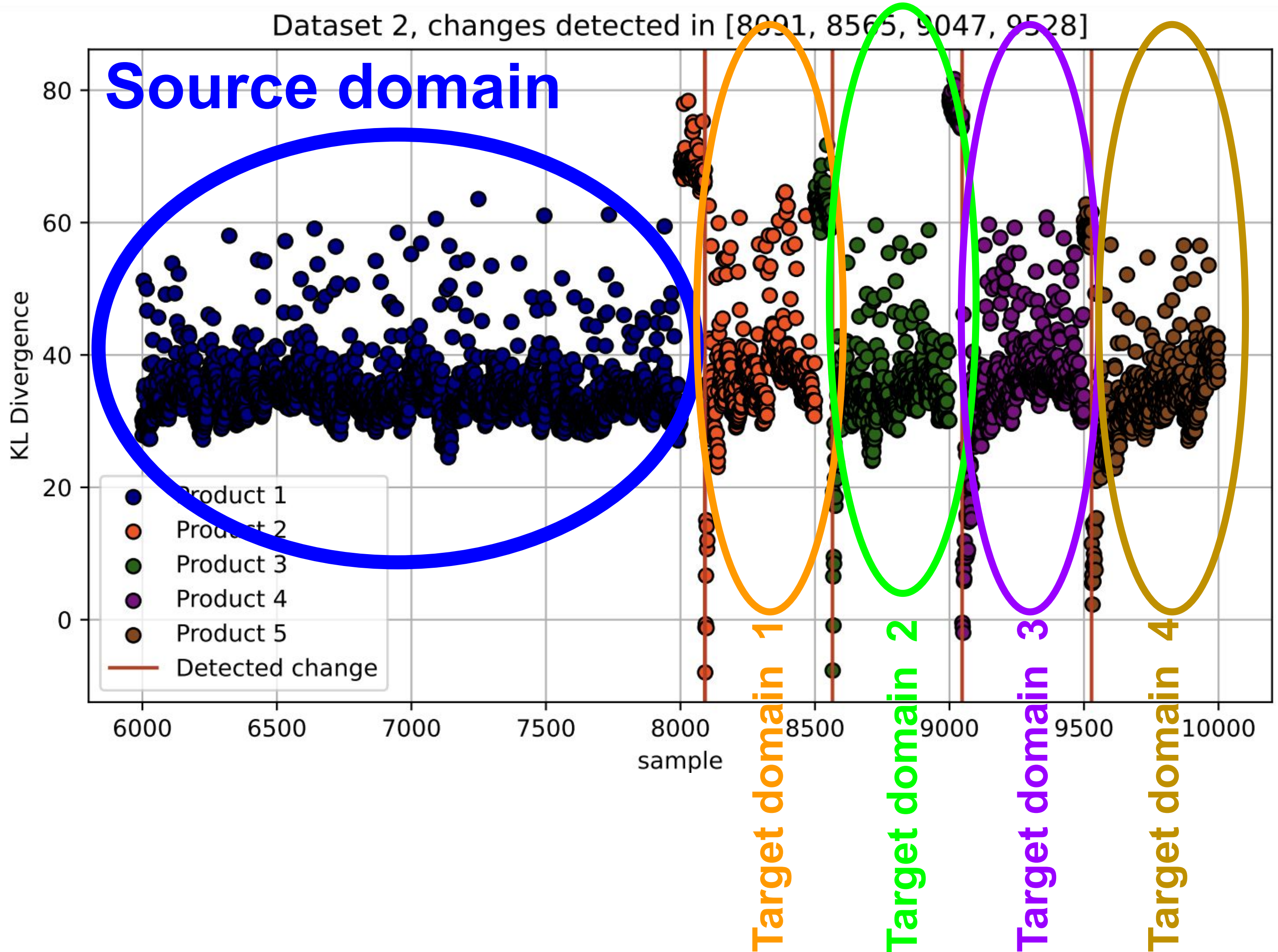


DOMAIN SHIFT DETECTOR
KL Divergence + Page Hinkley

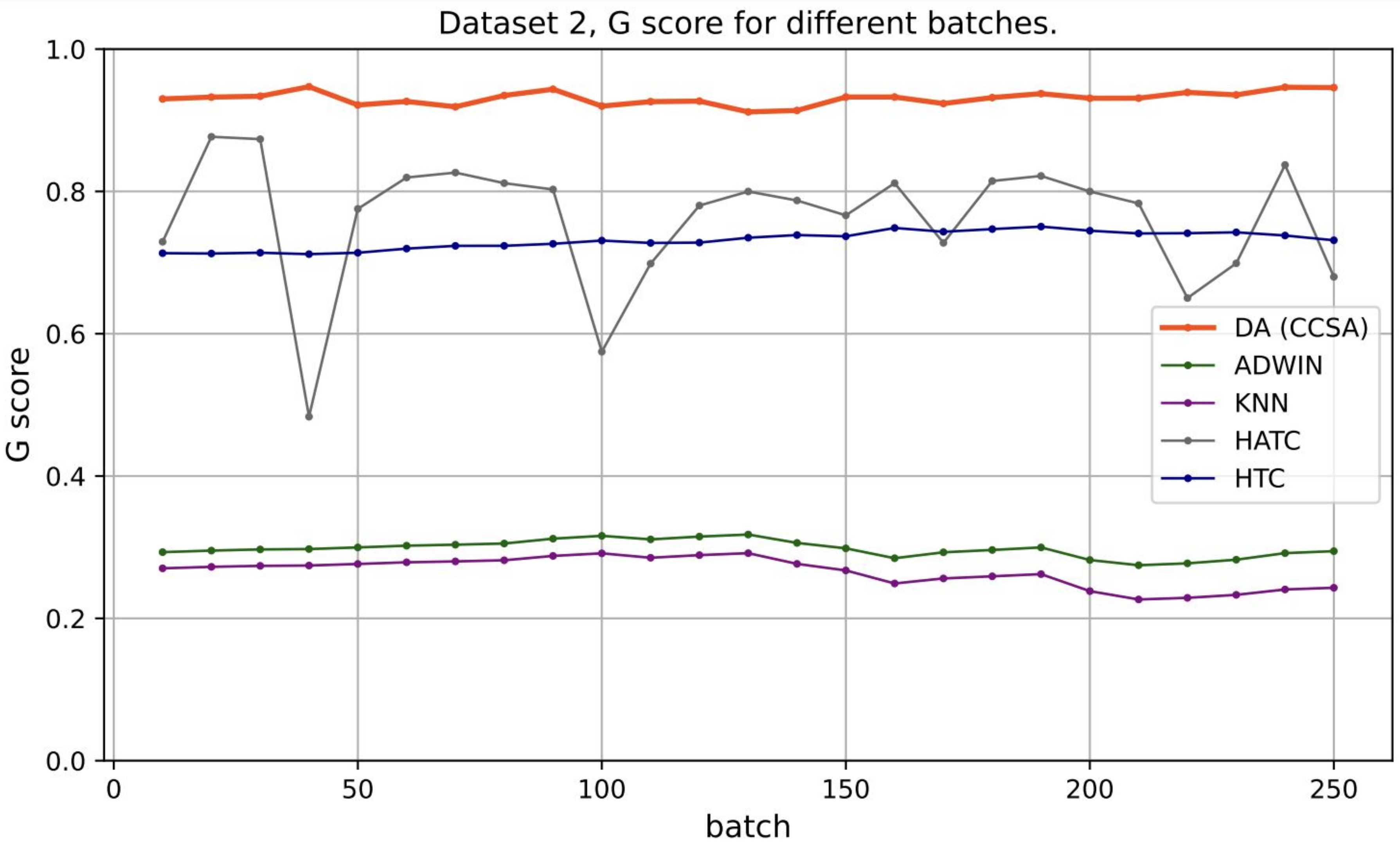


DOMAIN ADAPTATION CLASSIFIER

CCSA (Classification and Contrastive Semantic Alignment) Model



DOMAIN ADAPTATION CLASSIFIER

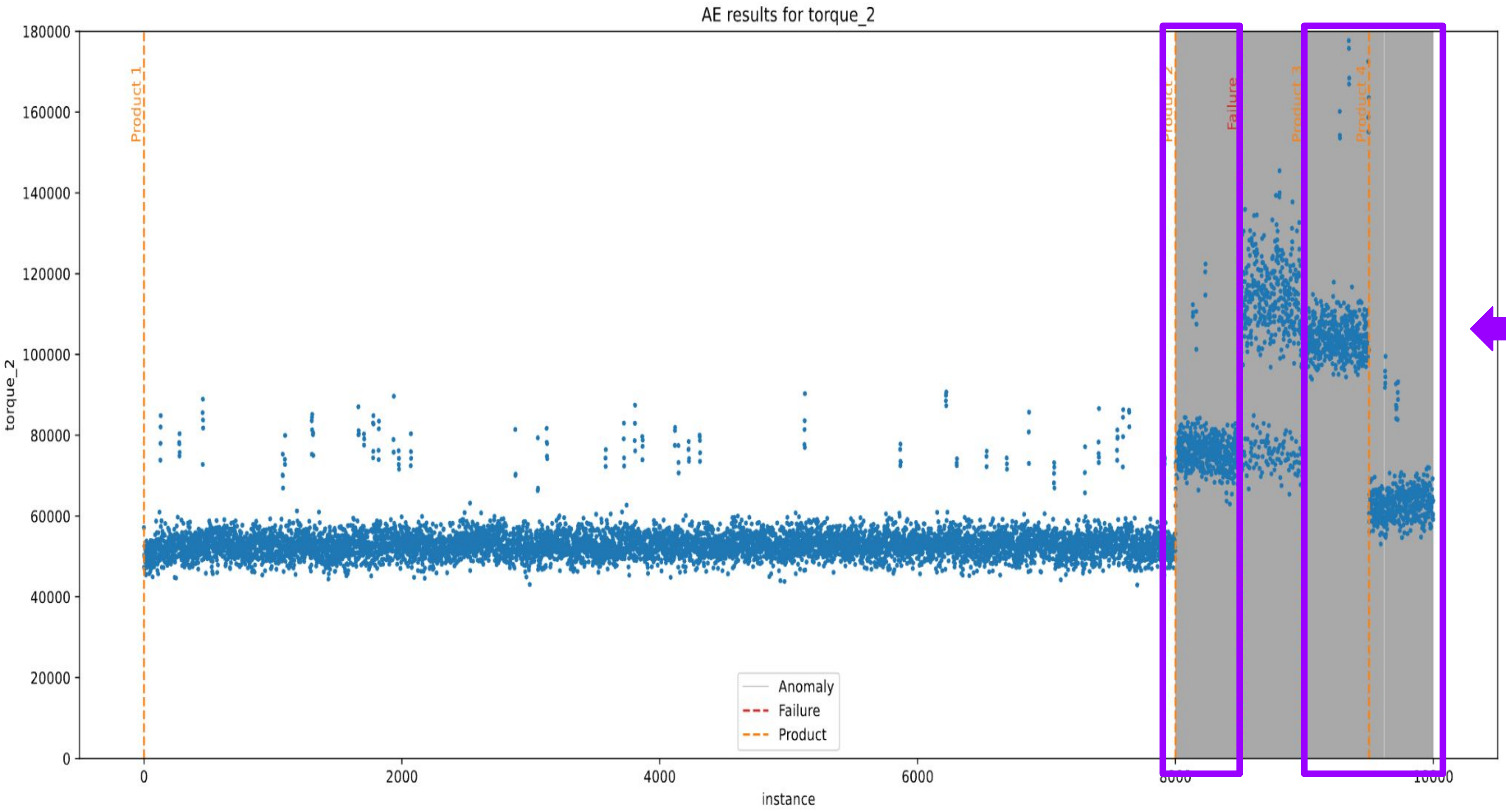


Towards Differentiating Between **Failures** and **Domain Shifts** in Industrial Data Streams

Natalia Wojak-Strzelecka, Szymon Bobek, Grzegorz J. Nalepa, Jerzy Stefanowski

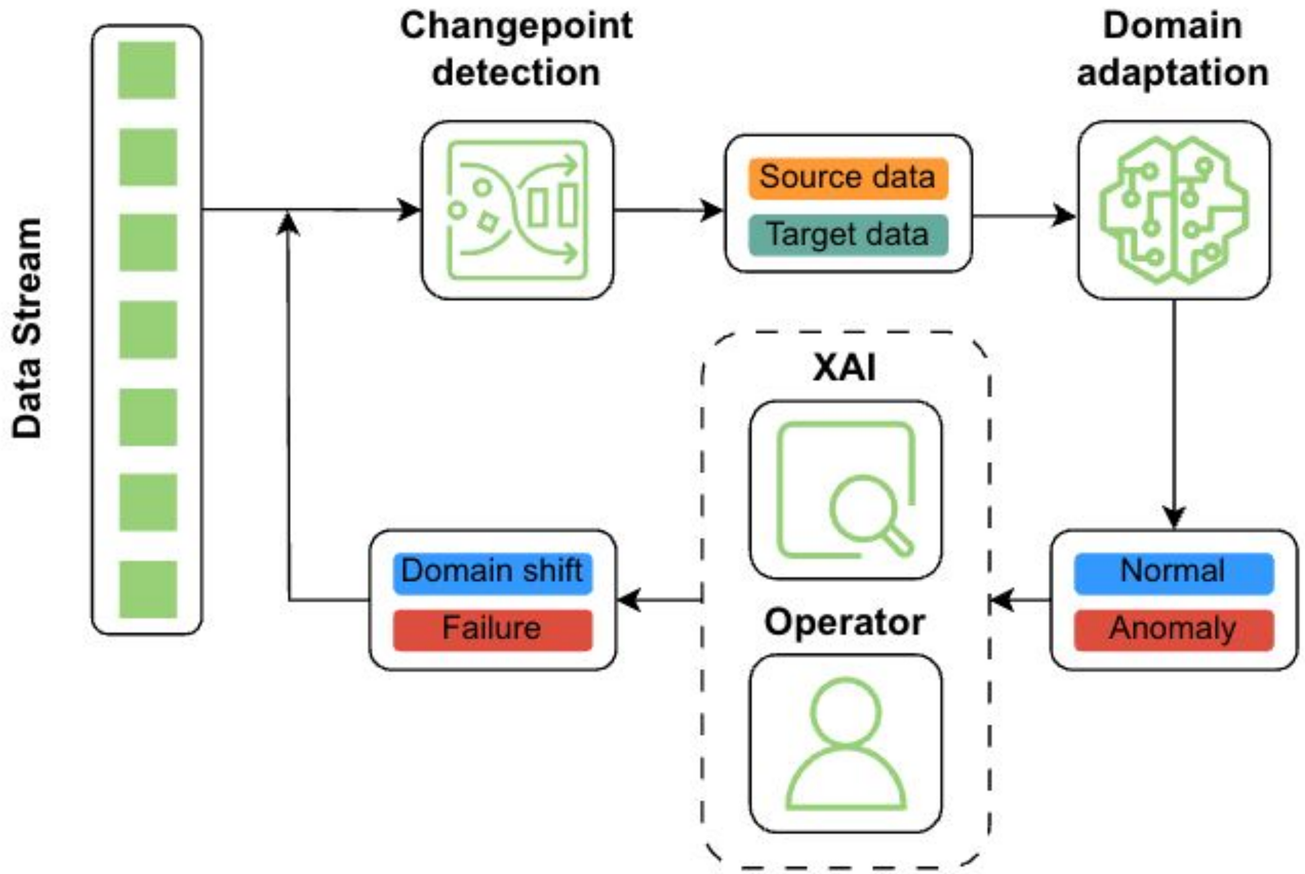
published on ECAI 2024
https://ruj.uj.edu.pl/entities/publication/ad84e59c-9427-46bd-b9a5-e70371bd2f77

DOMAIN SHIFT VS FAILURE



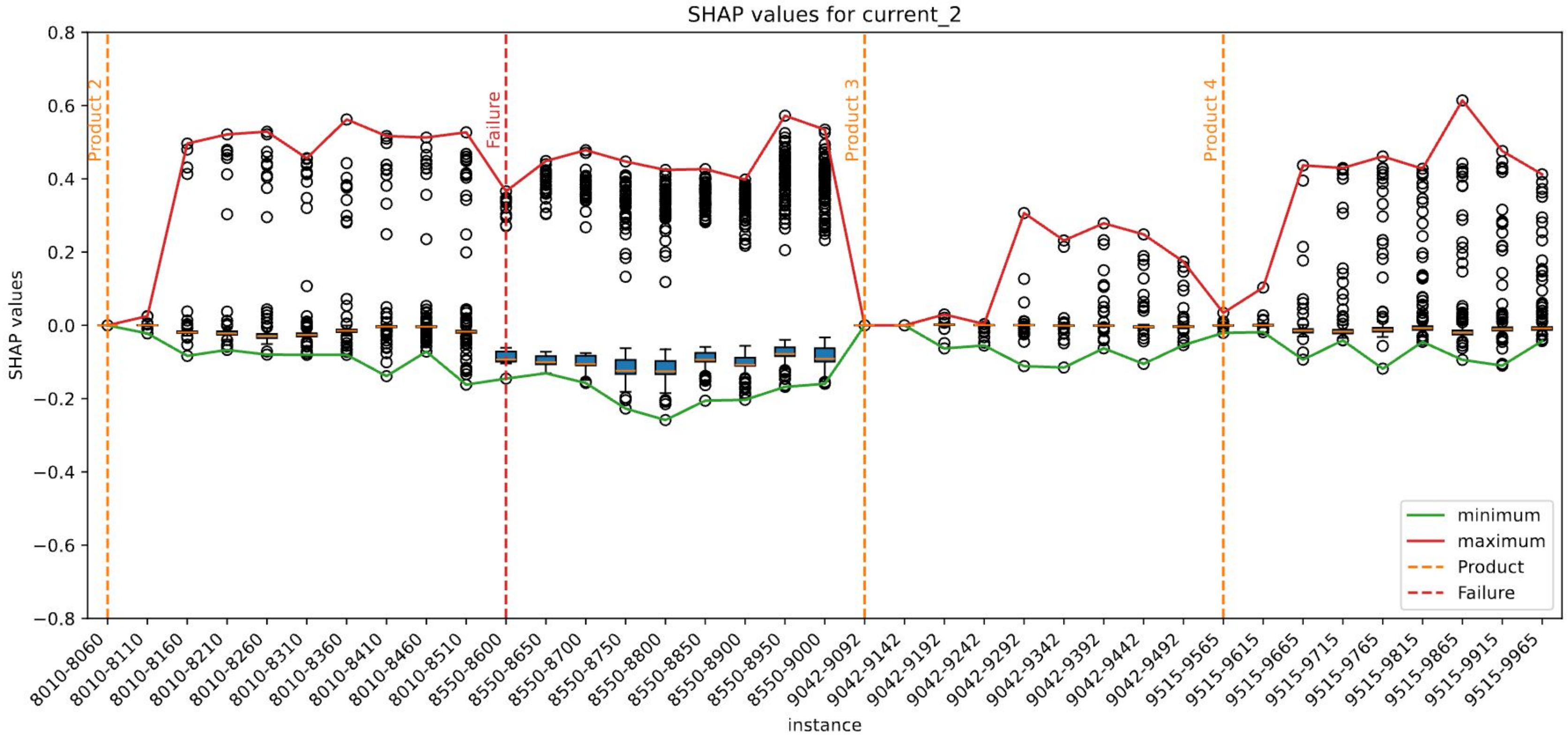
False alarms

PROPOSED APPROACH



GEIST
Research Group

EXPLANATION OF THE CHANGES OF MODEL BEHAVIOR BETWEEN CHANGEPOINTS



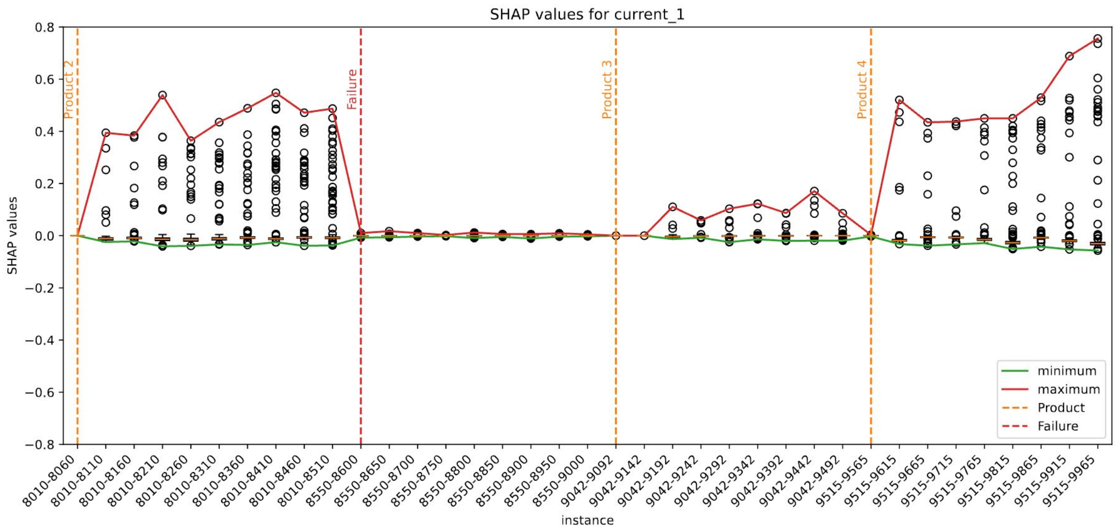
COLD ROLLING MILL DATASET

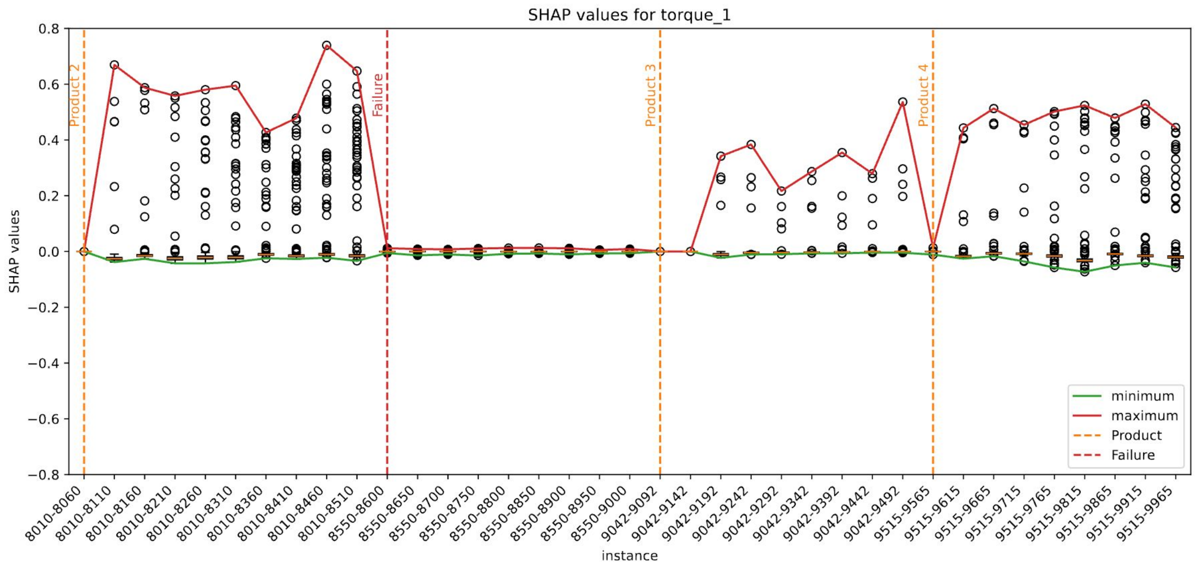
RESULTS

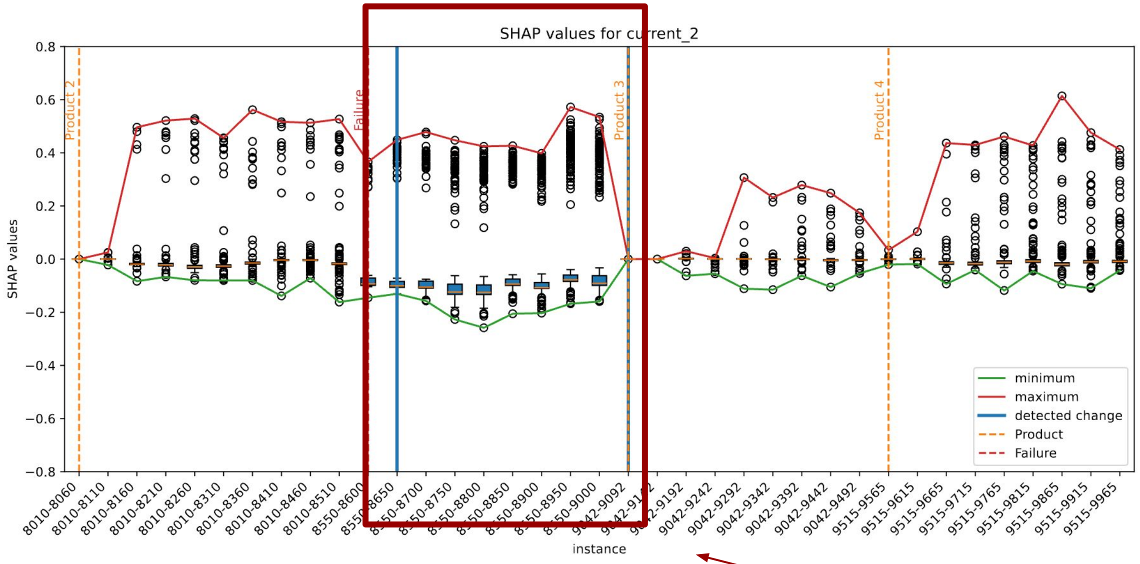
COLD ROLLING MILL DATASET

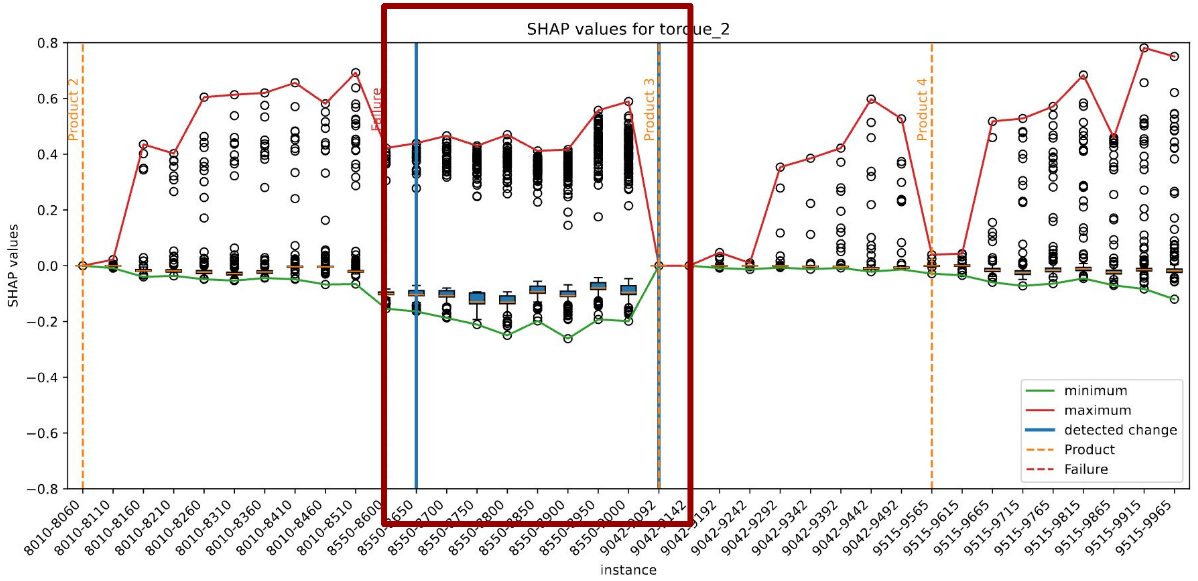
current

torque





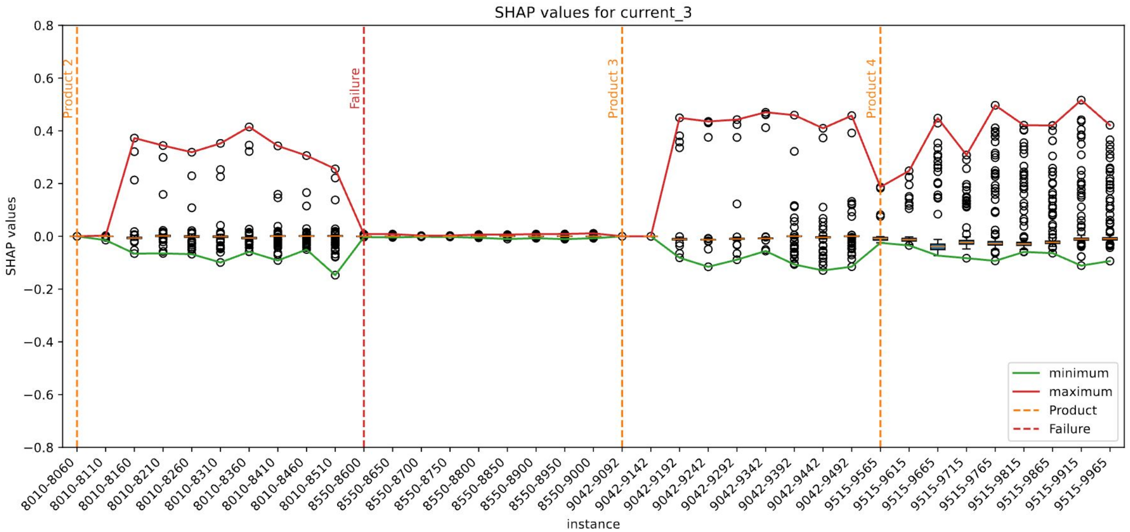


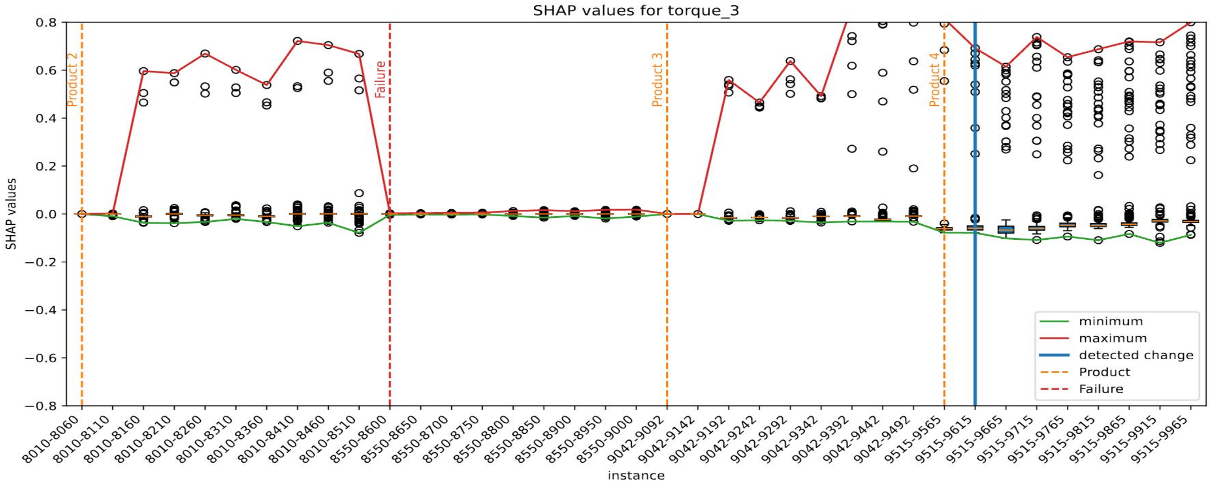


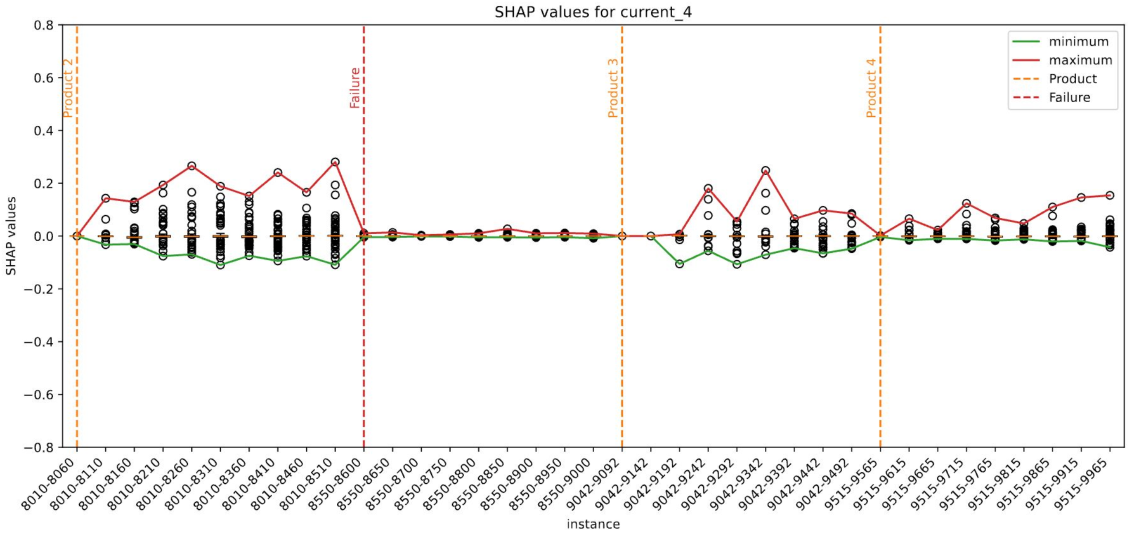
Failure

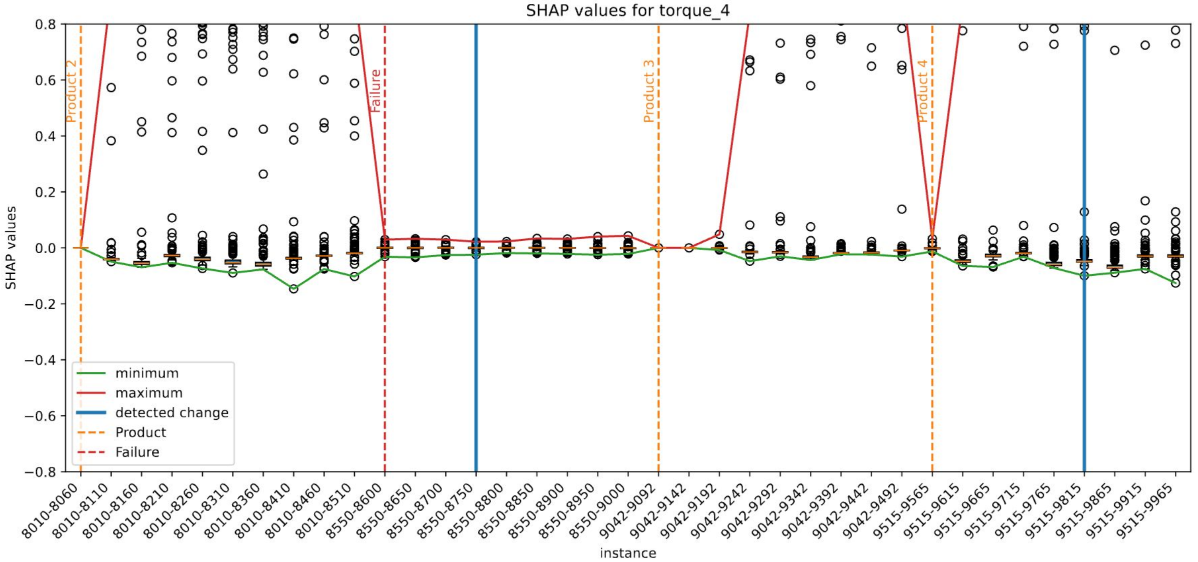
current

torque







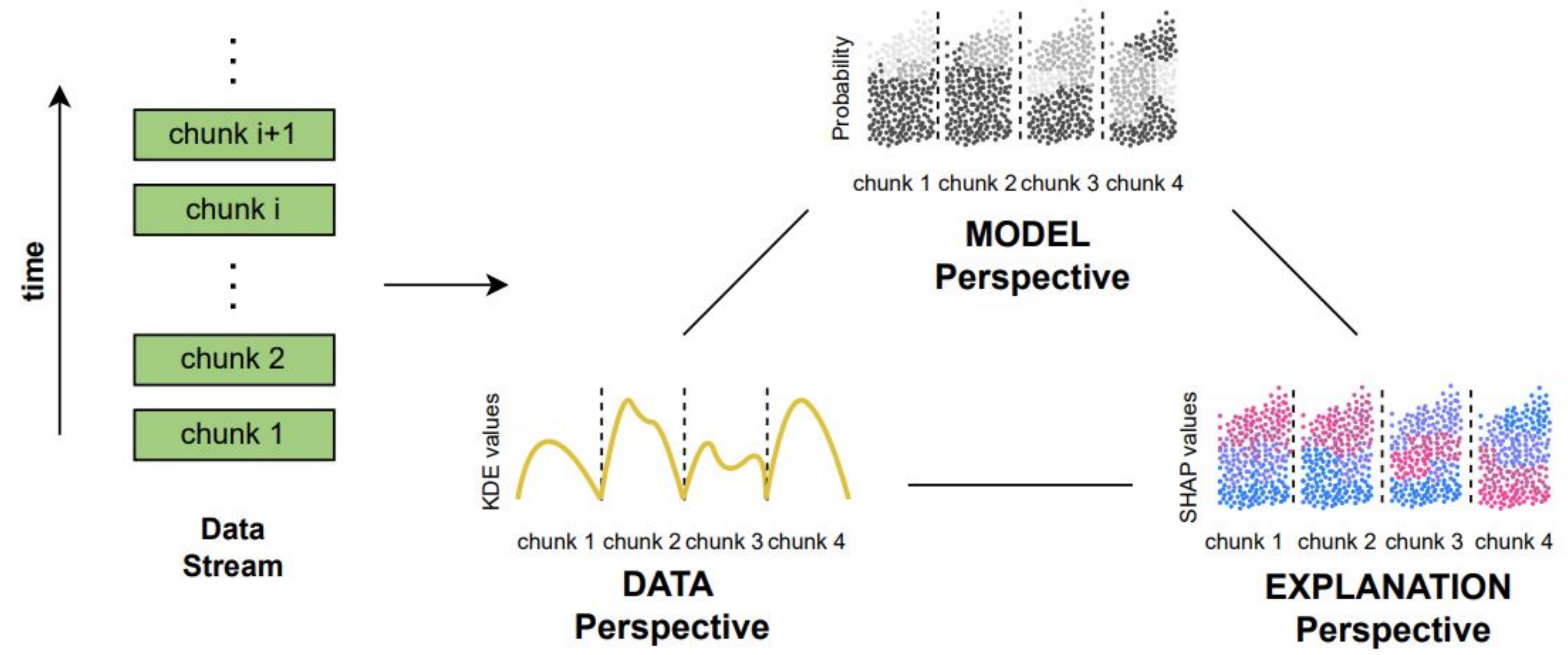


Proposing **multi-perspective approach** for detecting and explaining concepts drifts in evolving data

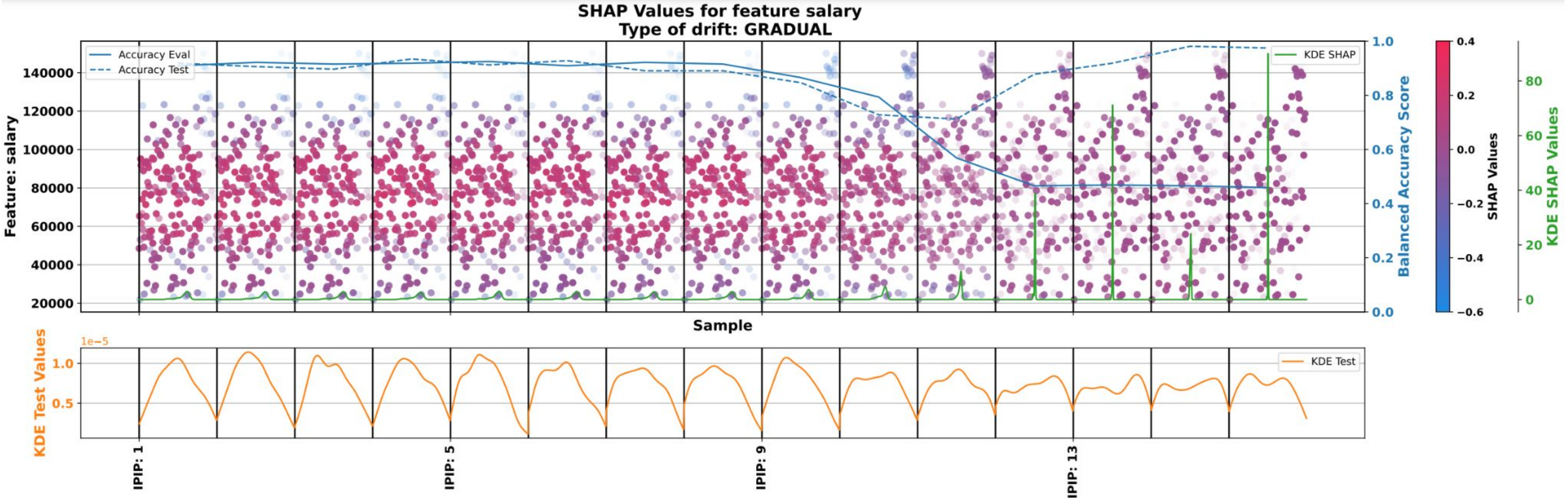
Natalia Wojak-Strzelecka, Antonio Guillén-Teruel, Szymon Bobek, Grzegorz J. Nalepa, José Palma, Juan A. Botía

submitted to ECML2025 research track

FLOWCHART of MPF (multi-perspective framework)

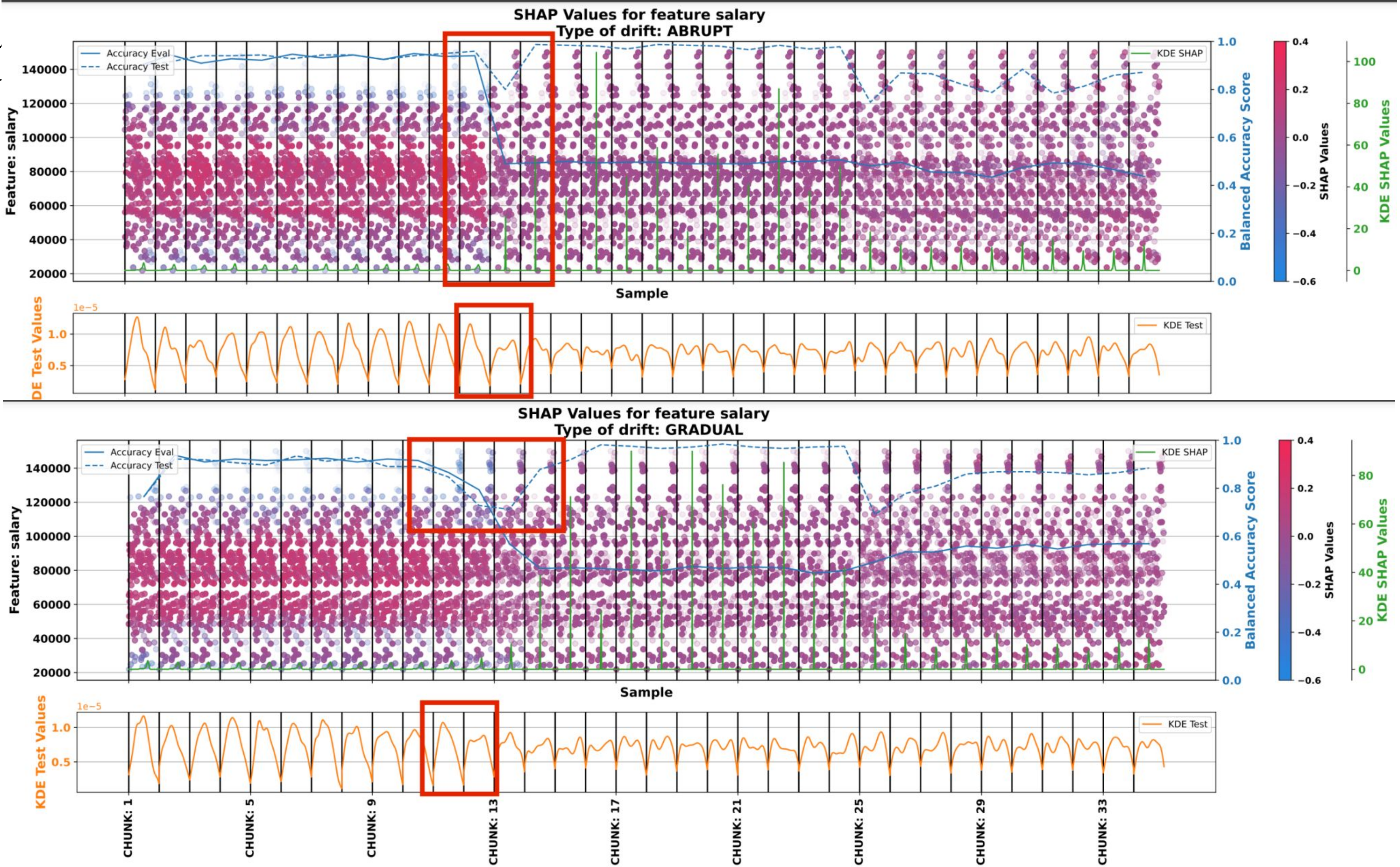


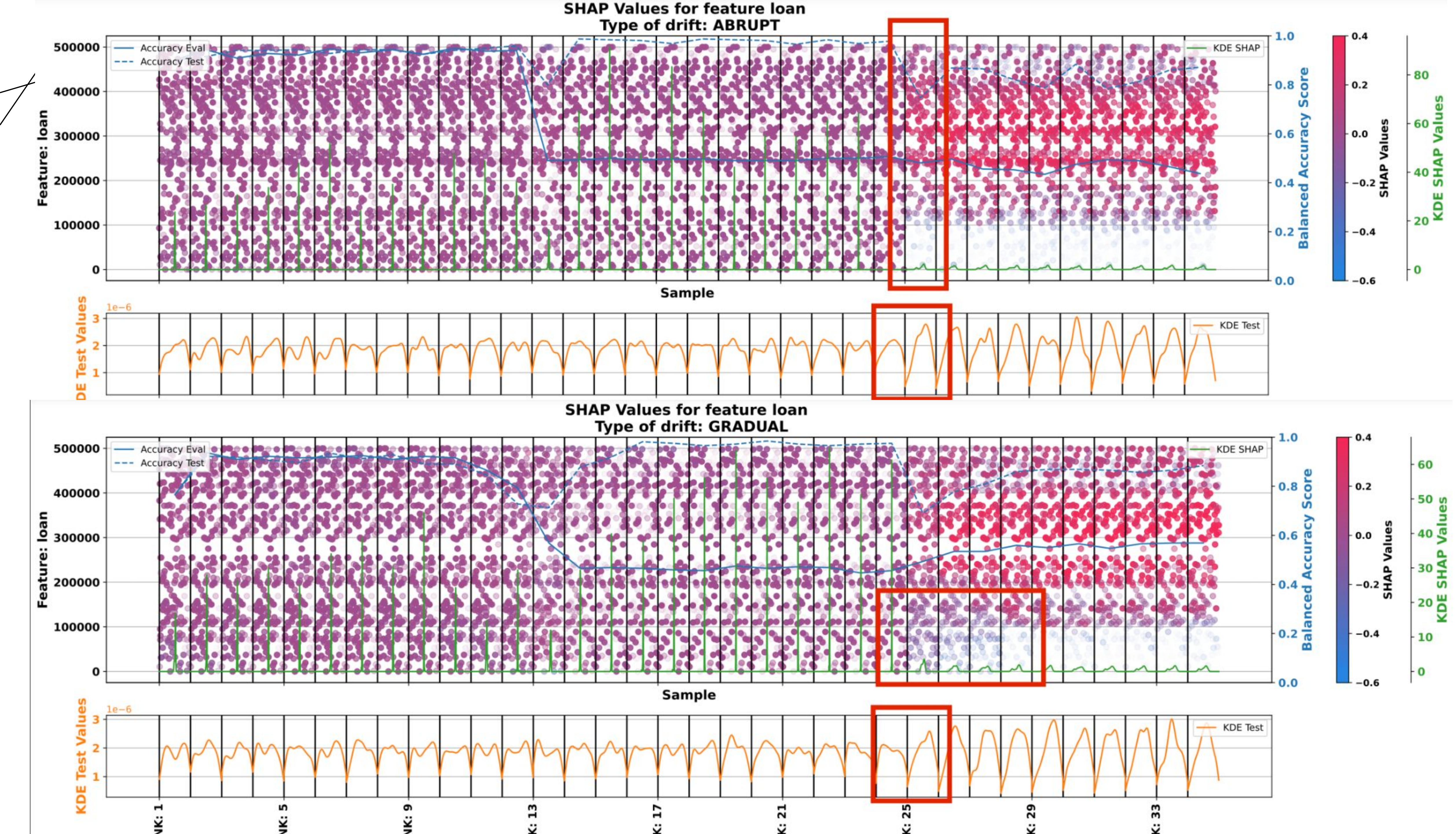
EXAMPLE OF VISUALIZATION



RESULTS
AGRAWAL DATASET

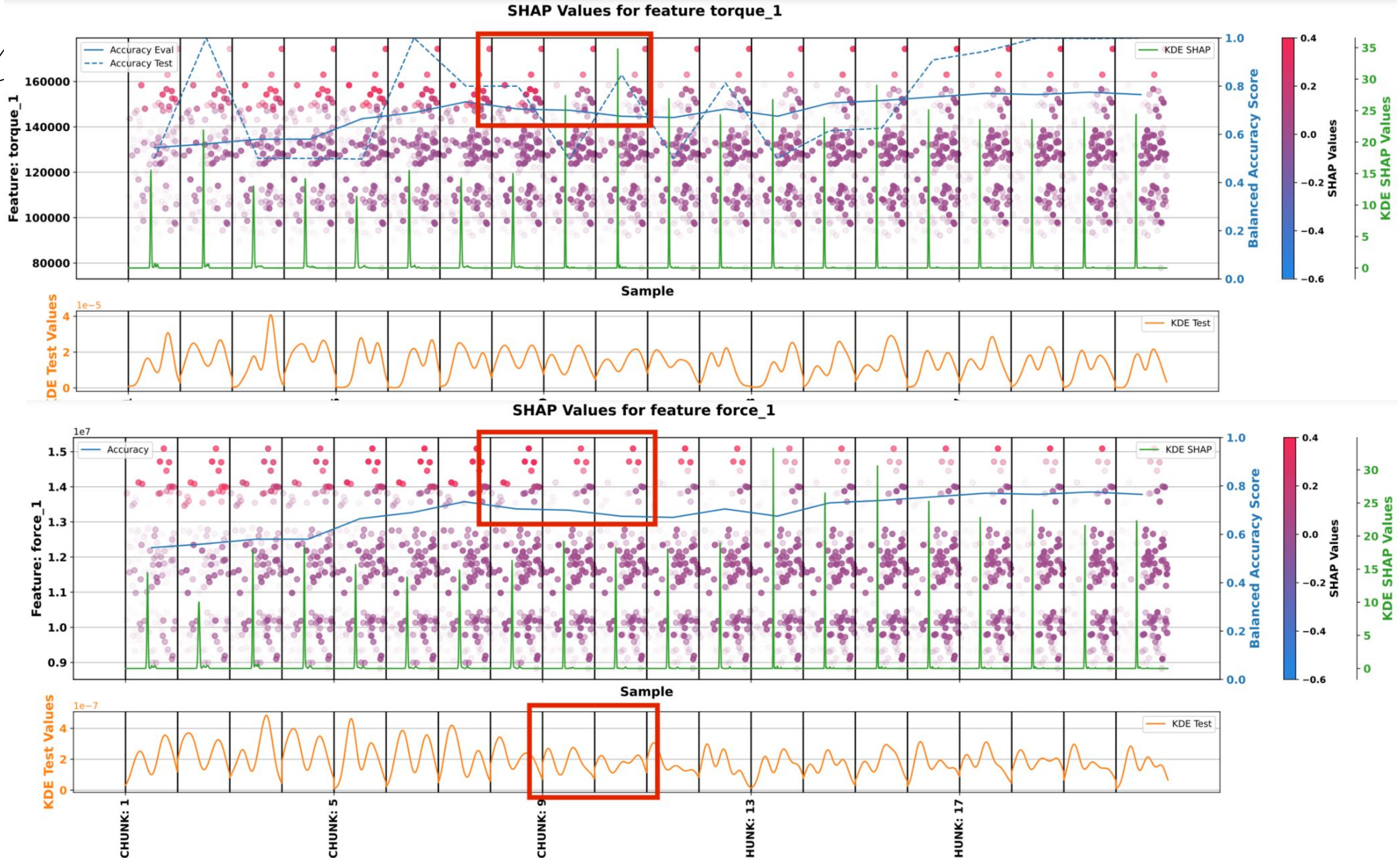
- Data stream with nine features, including six numerical and three categorical attributes.
- Assign binary class labels, likely indicating whether a loan should be approved.
- Data stream divide into 34 chunks.
- Concept **drift around 12 and 24 chunk.**

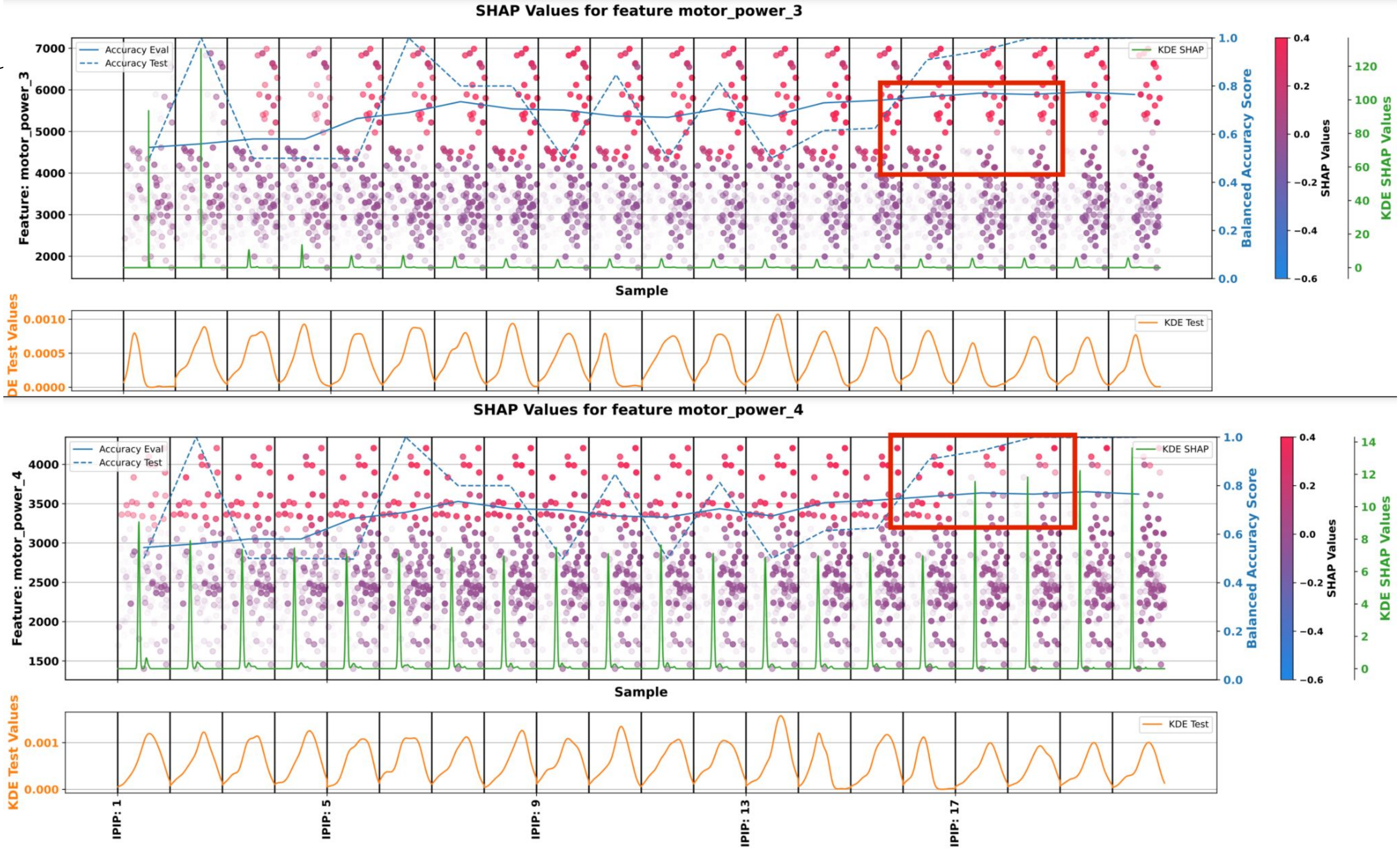

SHAP Values for feature salary
Type of drift: ABRUPT
Accuracy Eval
Accuracy Test
KDE SHAP
Feature: salary
Balanced Accuracy Score
SHAP Values
KDE SHAP Values
Sample
KDE Test Values
KDE Test
SHAP Values for feature salary
Type of drift: GRADUAL
CHUNK: 1
CHUNK: 5
CHUNK: 9
CHUNK: 13
CHUNK: 17
CHUNK: 21
CHUNK: 25
CHUNK: 29
CHUNK: 33


SHAP Values for feature loan
Type of drift: ABRUPT
Accuracy Eval
Accuracy Test
KDE SHAP
Feature: loan
Sample
Balanced Accuracy Score
SHAP Values
KDE SHAP Values
KDE Test
KDE Test Values
1e−6
SHAP Values for feature loan
Type of drift: GRADUAL
Accuracy Eval
Accuracy Test
KDE SHAP
Feature: loan
Sample
Balanced Accuracy Score
SHAP Values
KDE SHAP Values
KDE Test
KDE Test Values
1e−6
NK: 1
NK: 5
NK: 9
K: 13
K: 17
K: 21
K: 25
K: 29
K: 33

RESULTS
COLD ROLLING MILL DATASET

- Five–stand cold rolling mill simulator.
- Binary labels indicating normal or anomalous behavior.
- The dataset was divided into 20 equally sized chunks.
- **Two drifts:**
    - First, a **drift associated with the introduction of previously unseen product types** (around chunk 8).
    - Second, a **drift related to increased production capacity** is introduced (around chunk 16).

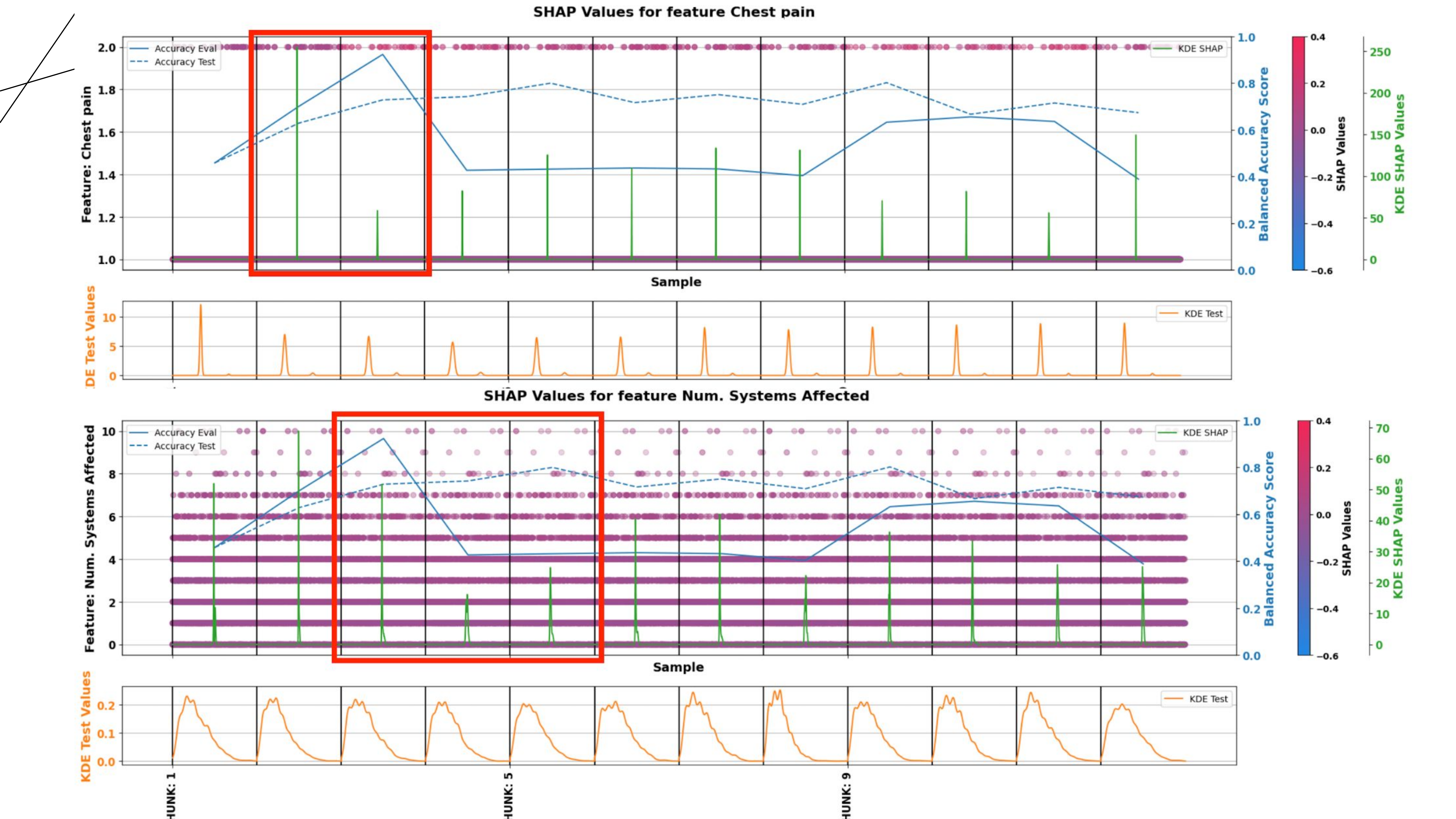

SHAP Values for feature torque_1
Accuracy Eval
Accuracy Test
KDE SHAP
Feature: torque_1
Sample
Balanced Accuracy Score
SHAP Values
KDE SHAP Values
KDE Test
KDE Test Values
SHAP Values for feature force_1
Accuracy
Feature: force_1
CHUNK: 1
CHUNK: 5
CHUNK: 9
HUNK: 13
HUNK: 17


SHAP Values for feature motor_power_3
Accuracy Eval
Accuracy Test
KDE SHAP
Feature: motor_power_3
Balanced Accuracy Score
SHAP Values
KDE SHAP Values
Sample
KDE Test
KDE Test Values
SHAP Values for feature motor_power_4
Feature: motor_power_4
IPIP: 1
IPIP: 5
IPIP: 9
IPIP: 13
IPIP: 17

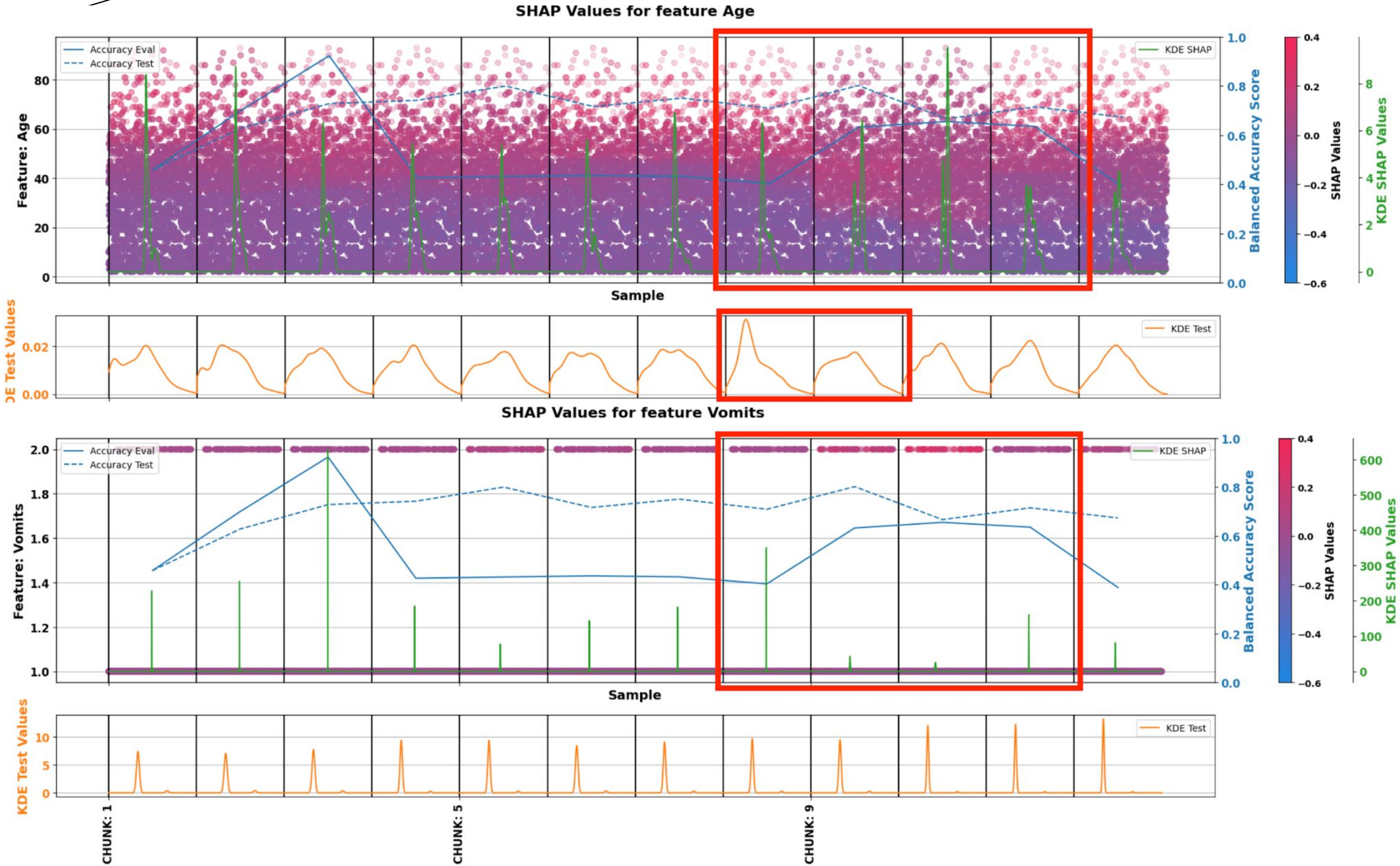
RESULTS

SMS COVID-19 DATASET

- These data were collected from electronic medical records of the Regional Health System (SMS) from the Region of Murcia (Spain).
- Binary classification problem - if a patient would need to be hospitalized or not .
- The dataset was divided into 13 equally sized chunks, each 2 months of data, starting from March 2020.
- **Two drifts:**
  - First, connected with the **variables responsible for hospitalization** (around chunk 2).
  - Second, connected with **changes affected by vaccination and different COVID variant** (around chunk 9).



SHAP Values for feature Chest pain
Accuracy Eval
Accuracy Test
KDE SHAP
Feature: Chest pain
Balanced Accuracy Score
SHAP Values
KDE SHAP Values
Sample
KDE Test
KDE Test Values
SHAP Values for feature Num. Systems Affected
Feature: Num. Systems Affected
CHUNK: 1
CHUNK: 5
CHUNK: 9


SHAP Values for feature Age
Accuracy Eval
Accuracy Test
KDE SHAP
Feature: Age
Sample
Balanced Accuracy Score
SHAP Values
KDE SHAP Values
KDE Test Values
KDE Test
SHAP Values for feature Vomits
Feature: Vomits
CHUNK: 1
CHUNK: 5
CHUNK: 9

FUTURE DIRECTIONS

UNSUPERVISED DOMAIN ADAPTATION

Extend the current supervised domain adaptation method to unsupervised or semi-supervised settings.

DRIFT TYPE DETECTION AND CLASSIFICATION

Introduce a module that automatically detect classifies the type of drift (e.g., sudden, gradual, recurring) based on SHAP trends and model performance indicators

ACTIVE LEARNING

Strategies to query for the most informative labels when drift is suspected, optimizing labeling cost.

HUMAN IN THE LOOP

User-friendly tools for non-experts to visualize the drift

Thank you for attention!

MEET THE TEAM
https://www.geist.re/start