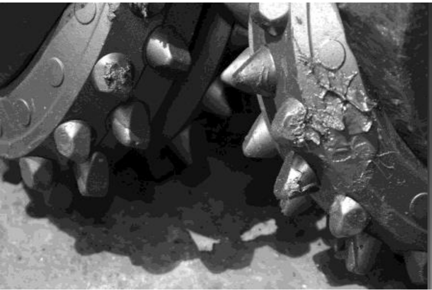
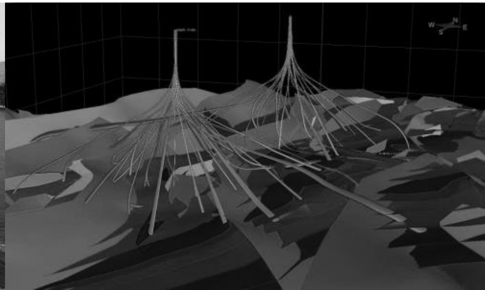
AI to improve data quality for automating drilling reports

Surface logging data automated QC

2024/10/03 - SPE Aberdeen





About Me

Brief bio

Fabio Concina

R&D Manager, kwantis

fabio.concina@kwantis.com

POLITECNICO MILANO 1863

BSc in Mathematical Engineering

UNIVERSITÀ DEGLI STUDI DI TRIESTE

MSc in Statistics

know anticipate succeed kwantis

In 2015 joined **kwantis** as a Software Engineer

In the last 9 years working on complex Software and Data Science problems

Now managing the R&D division

About kwantis

Innovative company with worldwide reach

Innovative company providing **Software** solutions and consulting to the Energy industry

Why Data Quality?

Why Data Quality?

Our use case

Our use case: **id3 Software**

id3 provides **advanced analytics for drilling operations**,
integrating HF and LF data for performance analysis and optimization

150+
Active drilling sites

35+
Countries

400+
KPIs

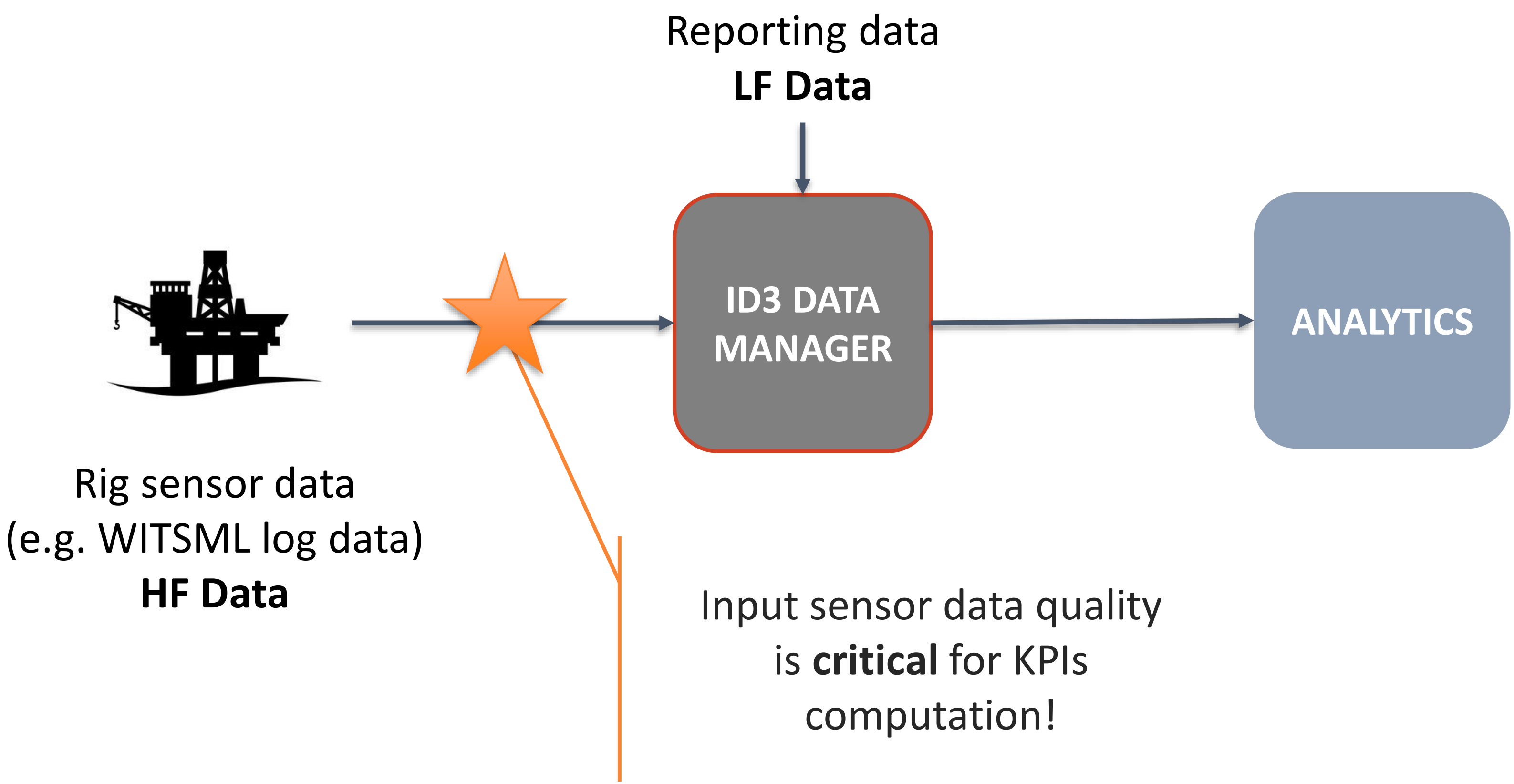
3000+
Processed wells

3B+
Records

id3 Integrated Drilling Data Discovery

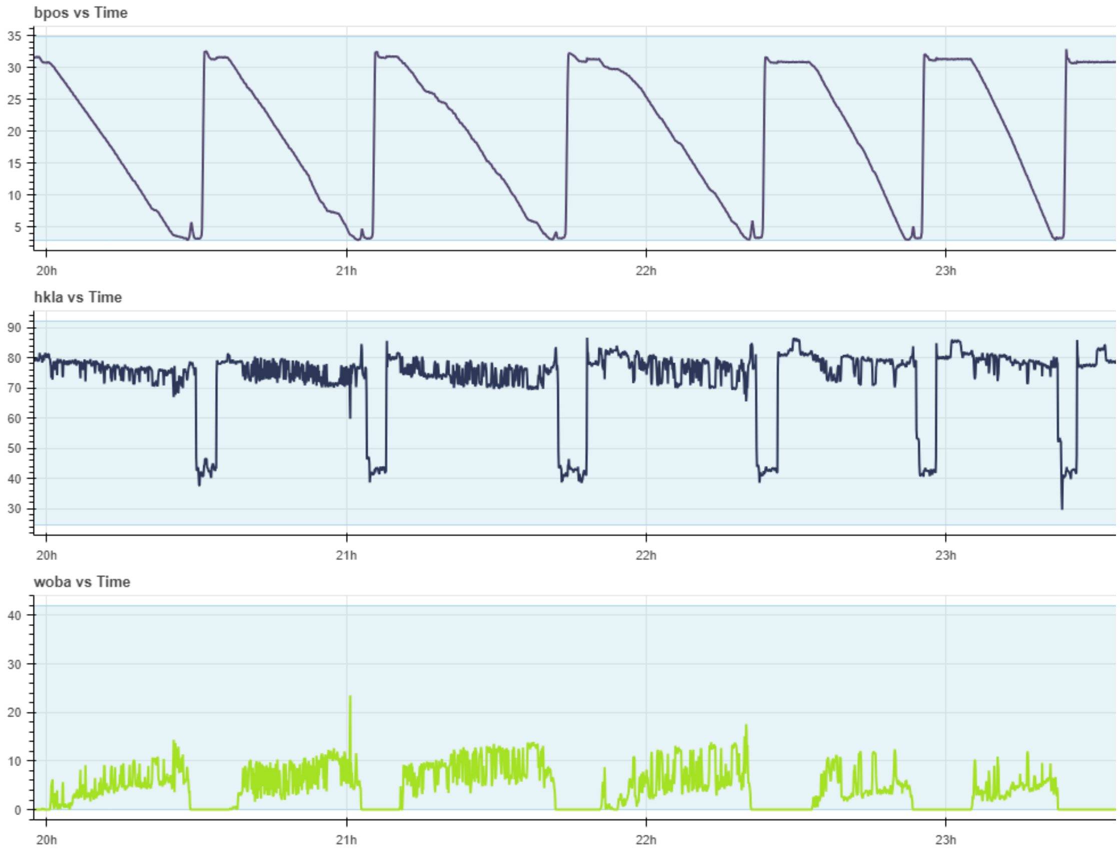
Why Data Quality?

Input sensor data quality



Why Data Quality?

Sensor data example



BLOCK POSITION

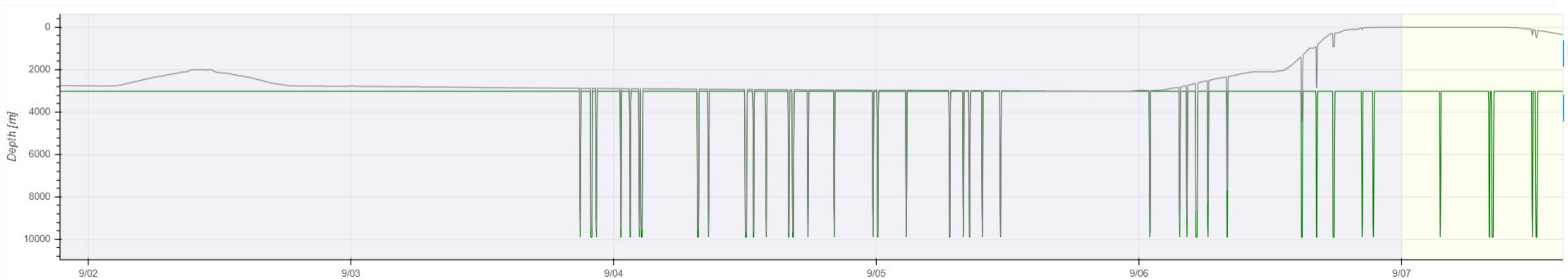
HOOK LOAD

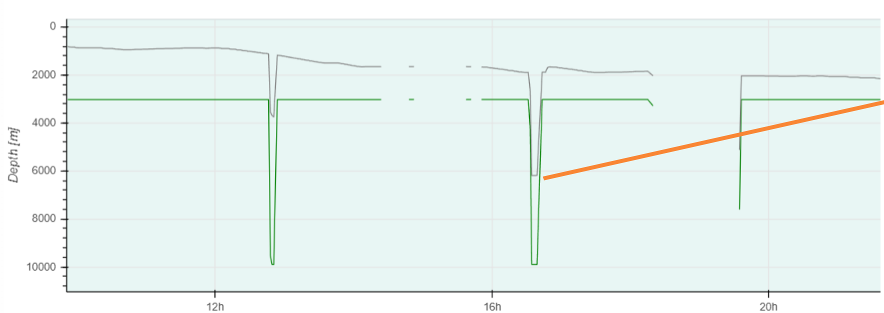
WEIGHT ON BIT

Why Data Quality?

Bad data quality

Examples of **bad** data quality



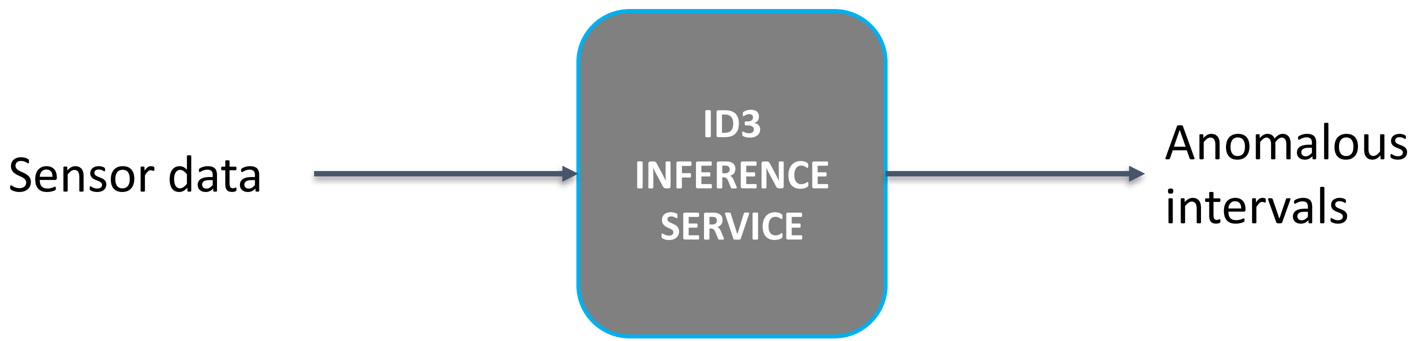


Spikes and **gaps** in depth data
(Bit position and Hole depth)

Our solution

id3 Inference Service

Our solution has been to **build a new id3 service** integrating traditional anomaly detection techniques and AI models



Our solution

2 Steps

Step 1

Univariate Anomaly Detection

Detecting anomalies using prior **expert knowledge**, for example:

- Value range, or value change range
- Intervals with constant value measurements
- Association rules between variables

Step 2

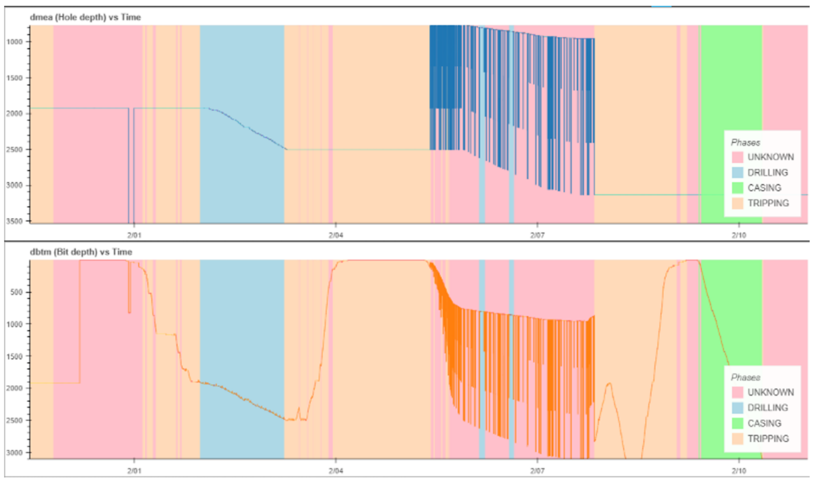
Multivariate Anomaly Detection

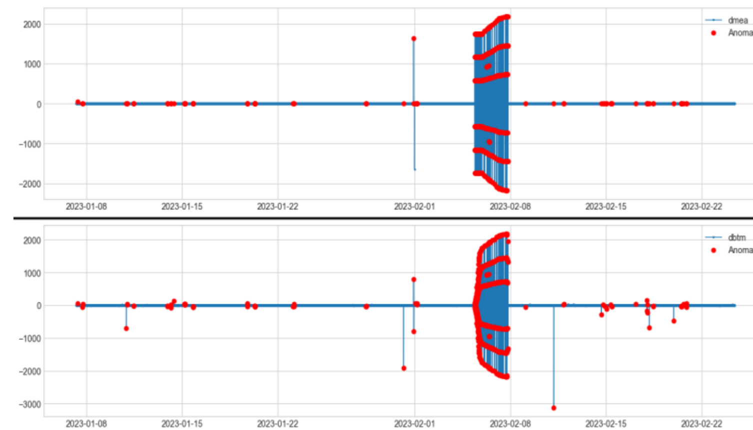
AI models considering several variables together. We are currently using an **autoencoder** architecture.

Univariate AD

Value range change

Example: **value range change**





Depth data exhibits incorrect oscillating behavior which is physically impossible

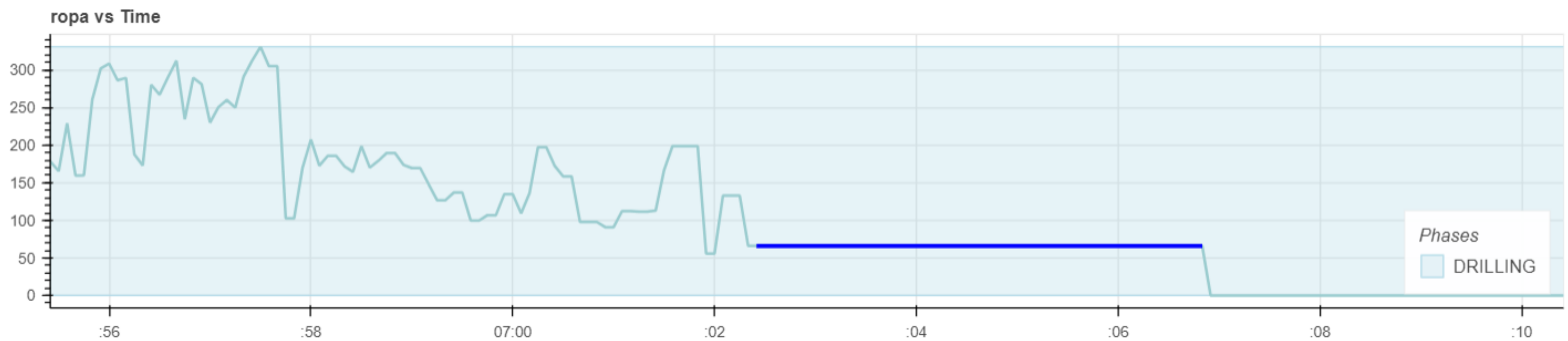
Differentiating the signal allows an easy detection of the anomalies

Univariate AD

Intervals with constant value measurements

Step 1

Example: **Intervals with constant value measurements**



In this case the Rate of Penetration (ROP) remains suspiciously **flat for several minutes**.
For different measurements is almost impossible to have the same value over an extended period

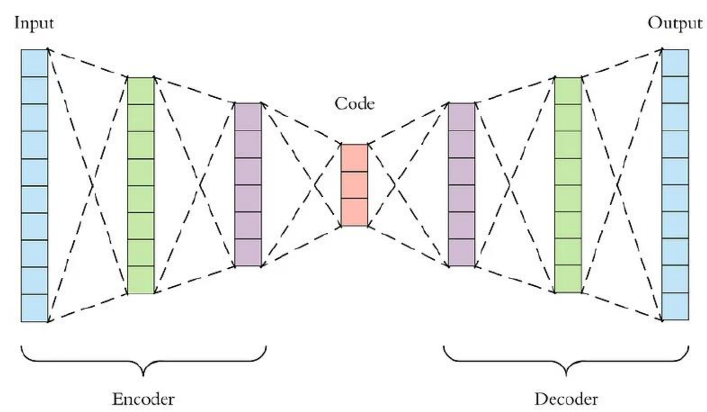
Multivariate AD

Autoencoder architecture

Step 2

Step 2 introduces AI methods to detect **more complex** anomalies, for example patterns not consistent with a normal drilling activity

We use the **autoencoder** neural network architecture, which is a type of artificial neural network designed for dimensionality reduction and feature learning

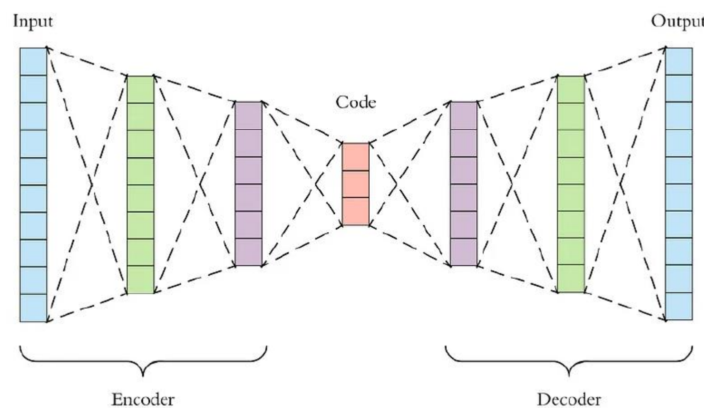


Multivariate AD

Autoencoder architecture

Step 2

The **input** to the network is a sliding window of the sensor data, e.g., 10 minutes

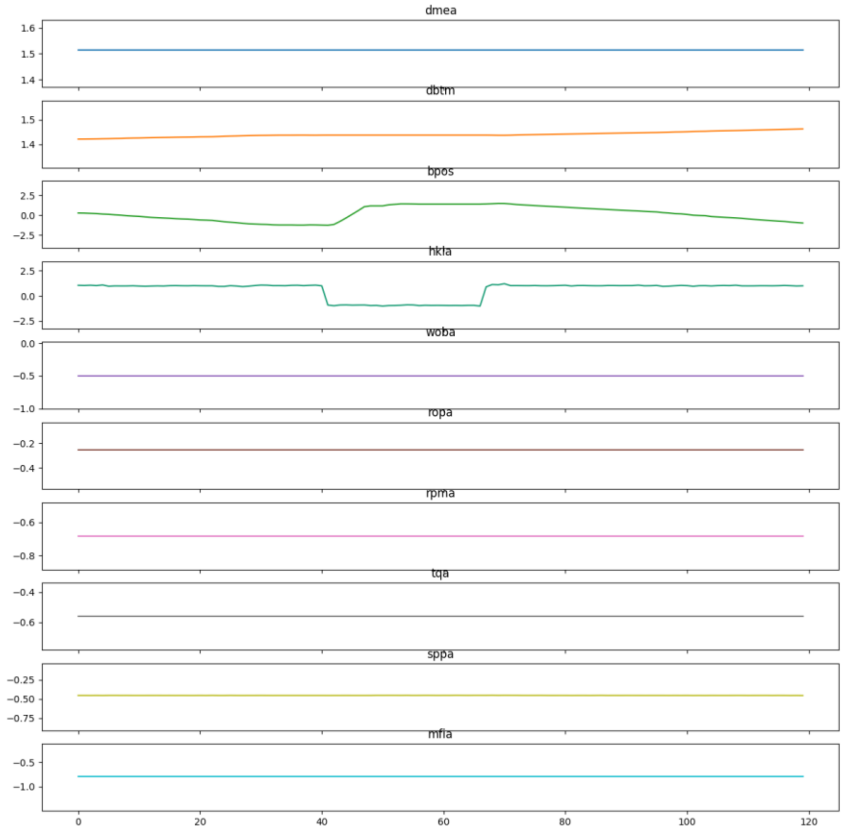


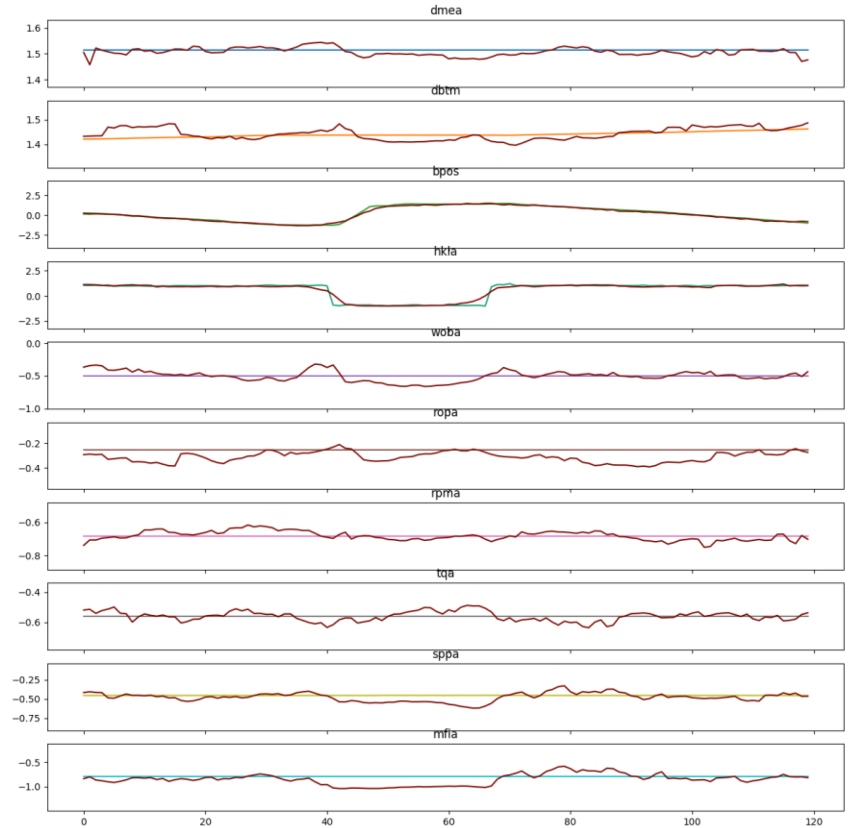
The **output** is again a window with the same size

The layer in the **middle** “compresses” the information and acts a feature extractor

Multivariate AD

Example input





Example of **input**

Example of **output**

Multivariate AD

Reconstruction error

Step 2

We define the **reconstruction error** for each data point as the squared difference of that data point and its reconstruction:

$$\mathrm{Error}(x_{i,t}, \hat{x}_{i,t}) = (x_{i,t} - \hat{x}_{i,t})^2,$$

Multivariate AD

Thresholding system

Step 2

The key idea is that normal signal is reconstructed with **low error**, whilst anomalous signal is reconstructed with **high error**

How can we say when error is **high**?

We need a **threshold**

Multivariate AD

Thresholding system

Step 2

Possible techniques for computing the **threshold**

- Percentile thresholding
- Bayesian Posterior Thresholding
- Z-score Thresholding

We devised and are using a *novel* thresholding technique based on class distribution, which requires labeled data

Multivariate AD

Example of a complex anomaly

Step 2

Example of
complex anomaly
detected

A drilling activity is ongoing, but during stand connection (when Hook Load drops) we have a **positive and high Weight on Bit**.

3 intervals are detected as anomalous

Future Work

Several lines of improvement

What’s next?

Currently training a **bigger, more powerful AI model**, with significantly more data

Transition to **density estimation** within the neural network architecture

Enhanced univariate anomaly detection procedures

Thank you