swisscom

Monitoring in the Big Data Era: to AD or not to AD?

Joana Soares Machado
Jelena Malić

29.03.2022

AMLD

Agenda

- Introduction
- Anomaly Detection in Practice
- Traditional Monitoring Methods
- Conclusion

"The fault is not in our stars..."
but where is it exactly?

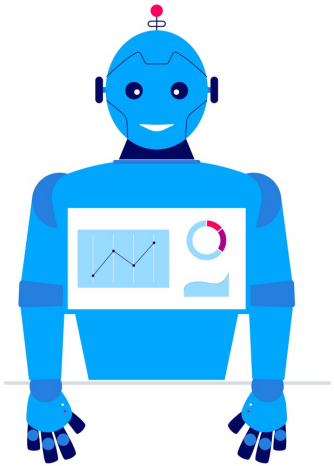
--Julius Caesar

Source of all Shakespeare images: stickers.cloud

Dilemma of the Modern Big Data Age

Apply Anomaly Detection?

- Use Machine Learning algorithms.



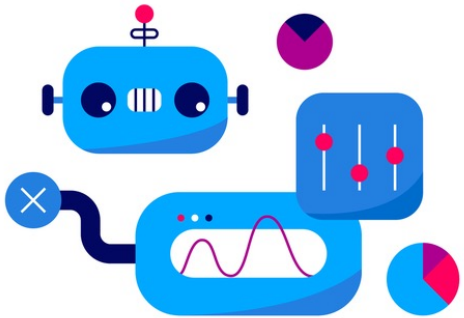
Apply Traditional Monitoring Methods?

- Use thresholds provided by experts.

Swisscom Use Cases

Network Infrastructure

20B interactions/day
on mobile network



Software Applications

300M function calls/day

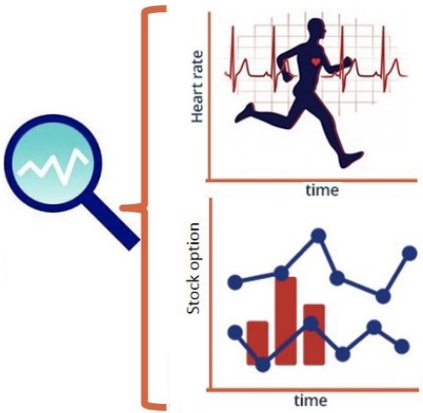
Business Processes

80k processes/day

Anomaly Detection Theory

Time Series

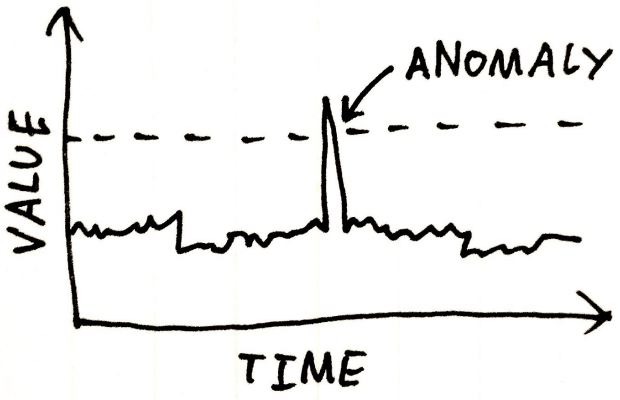
- Series of **data points ordered by time** of occurrence.



Source: Medium

Anomaly Detection

- Focused on identification of data points that are significantly **different from the majority of data.**



Source: Towards Data Science

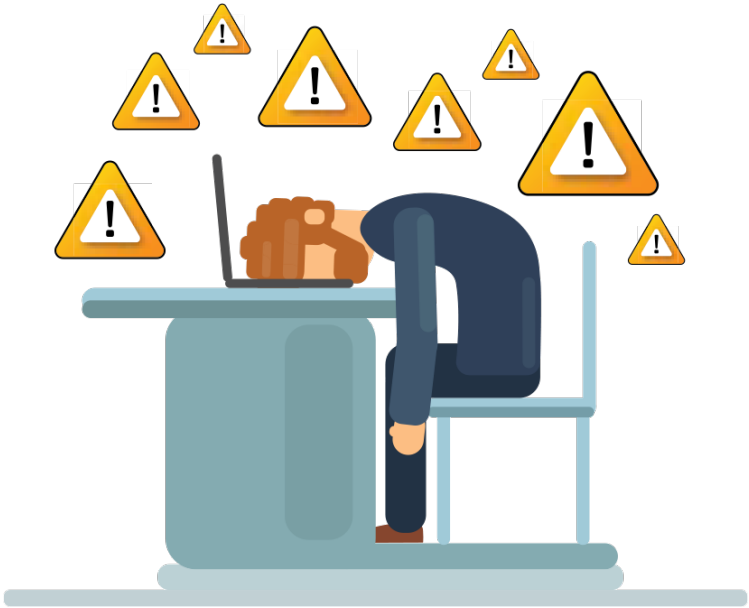
Anomaly Detection in Practice

“How far that little candle throws his beams!”

-- The Merchant of Venice

Anomaly Detection Challenges at Swisscom

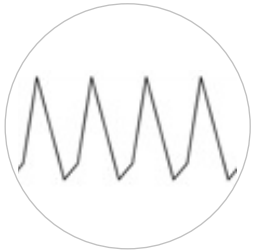
- **Efficiency**: 100,000+ time series in real-time.
- **Accuracy**: lack of labels.
- **Alerting**: false positives.
- **Generalization**: different data patterns.
- **Communication**: model interpretability.



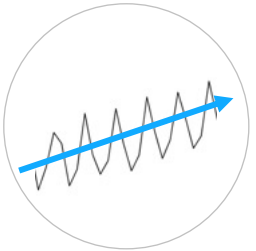
Source: Sensu

Anomaly Detection Approach

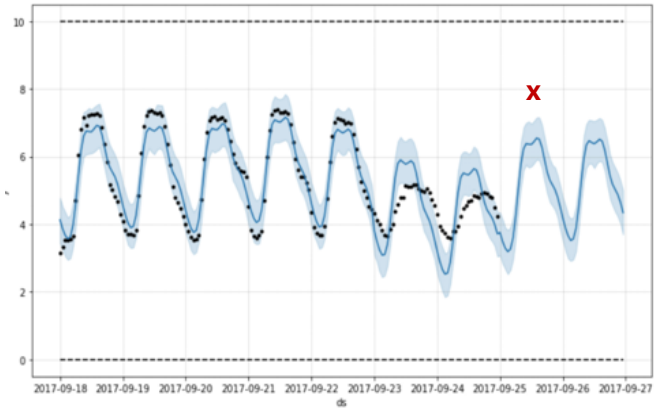
- **Unsupervised learning**: no labelled data.
- Detect **contextual** anomalies based on seasonality (daily, weekly) and trend.
- **Predictive models**: detect anomalies when an observation is outside of the prediction bounds.



Seasonality



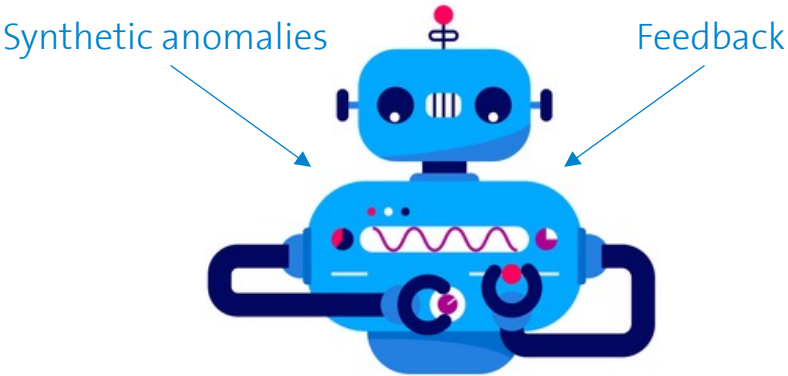
Trend

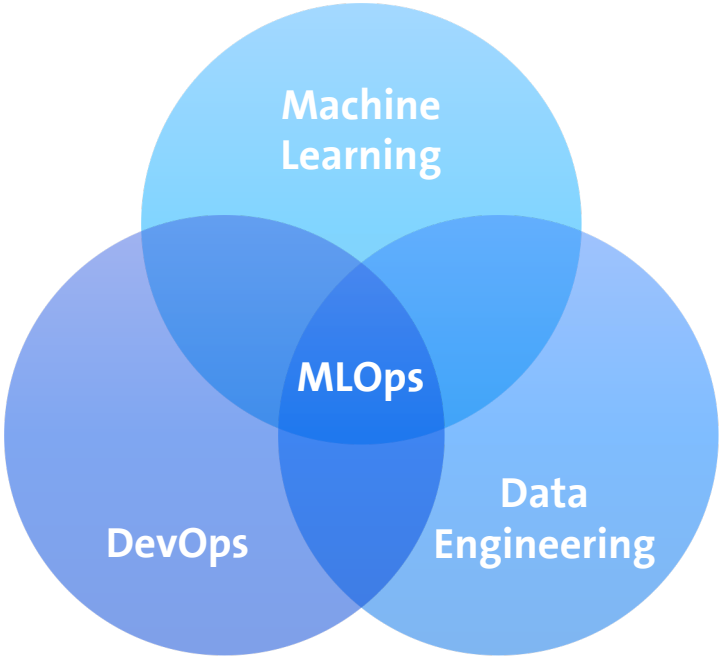


Open-source Model: *Prophet* [1].

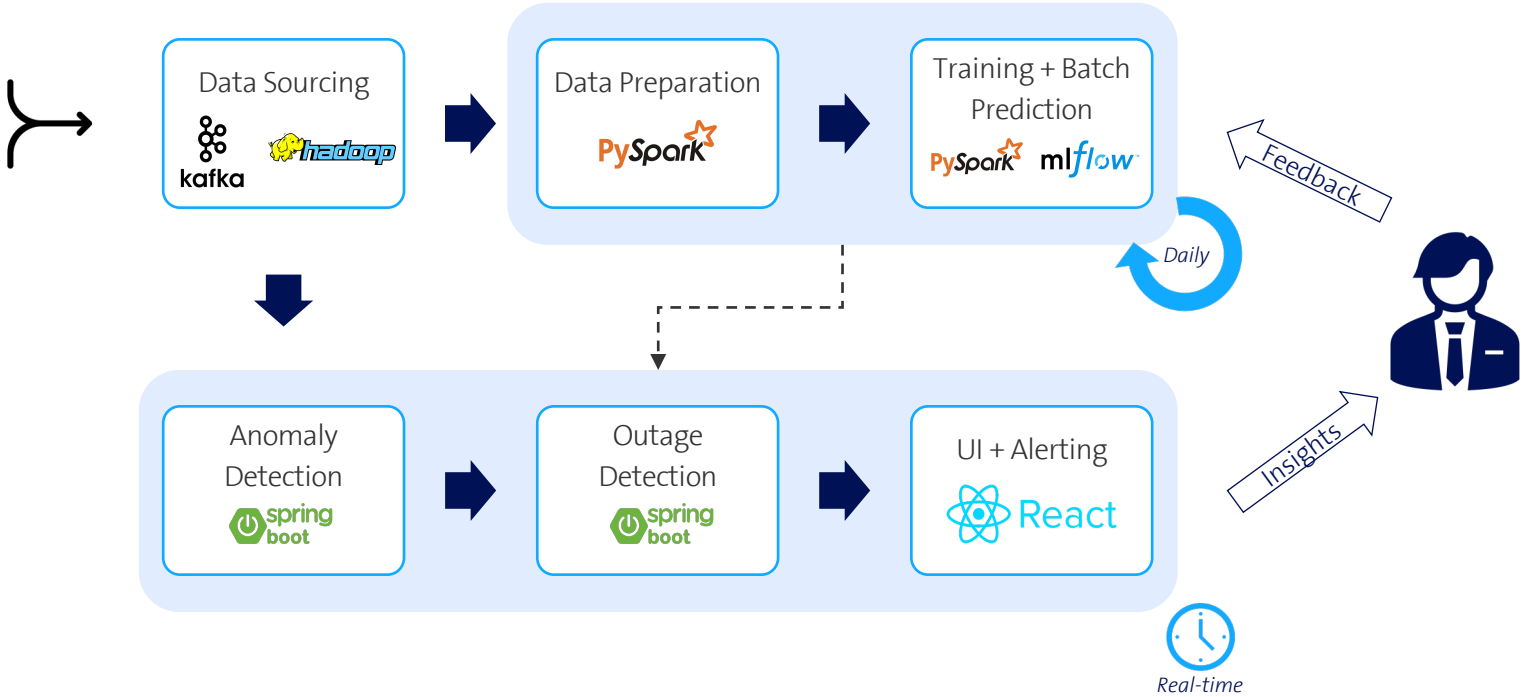
AutoML and MLOps

- Automatic **hyperparameter tuning**.
- Automatic **model selection**.



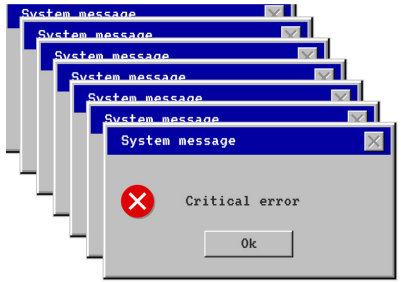


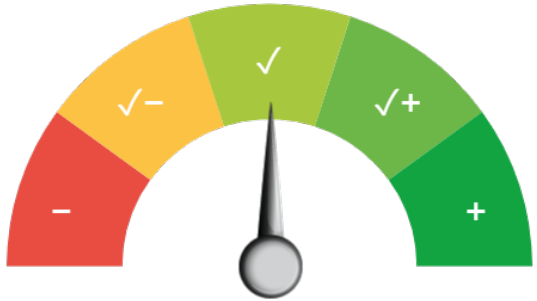
Machine Learning Pipeline



Monitoring with Anomaly Detection

- Consider **additional factors** for alerting:
  - Number of consecutive anomalies;
  - Outlier score;
  - Model maturity.
- Aggregate anomalies into higher-level concepts like **outages or health scores**.
- **Root cause analysis** is used to get actionable insights.





Traditional Monitoring Methods

“There is nothing either good or bad,
but thinking makes it so.”

-- Hamlet

Traditional Approach Challenges in Swisscom

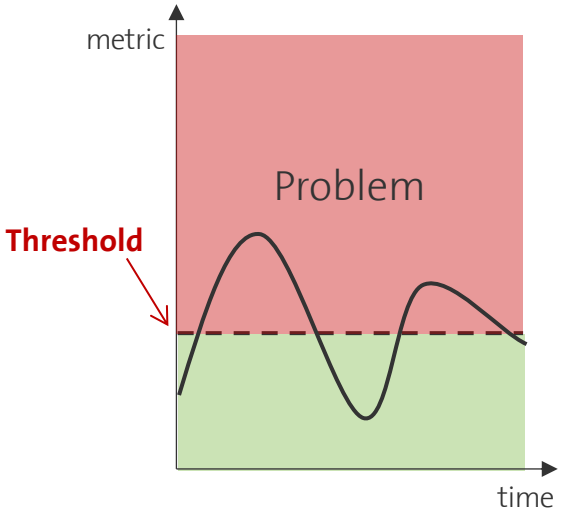
- **Efficiency**: 100,000+ time series in real-time.
- **Accuracy**: lack of labels.
- **Alerting**: false positives.
- **Generalization**: system evolution.
- **Communication**: requires domain experts.

Source: Freepik

Threshold-Based Approach

- **Heuristics** derived from **business knowledge**.
- **Rates instead of absolutes** make the approach more robust against data distribution changes!

- Examples:
  - ~~Alert if - *# Errors > 100 in 5 min*~~
  - Alert if - *Error Rate > 10% in 5 min*



SLI/SLO* Approach

- Based on *Site Reliability Engineering* book (2016).
- **Generic customer-focused approach** for monitoring of metrics.
- **SLI = Metric**

$$SLI = \frac{\textbf{good}\ events}{\textbf{valid}\ events} * 100\ (\%)$$

$$SLI = 1 - \frac{\textbf{bad}\ events}{\textbf{valid}\ events} * 100\ (\%)$$

- **SLO = Target (Threshold)**

Value should be within the target,
X% of the time window T

* SLI = Service Level Indicator, SLO = Service Level Objective

SLI/SLO Example – Ice Cream Shop

- **SLI** - Time taken by the shop to prepare an ice cream for a customer.
- **SLO** – In a month, 90% of the time customers get ice cream within 2 minutes.
- **Error Budget** – When an ice cream flavour runs out, it can take more than 2 minutes to serve the next customer (including the time to refill the ice cream bucket).

Source: VectorStock

Self-Service as a Solution

- Enables experts to:
  - **Interact** with their data;
  - **Express** their system expectations;
  - **Version** their monitoring setup;
  - **Communicate** with other stakeholders;
  - **Keep customer perspective central**.

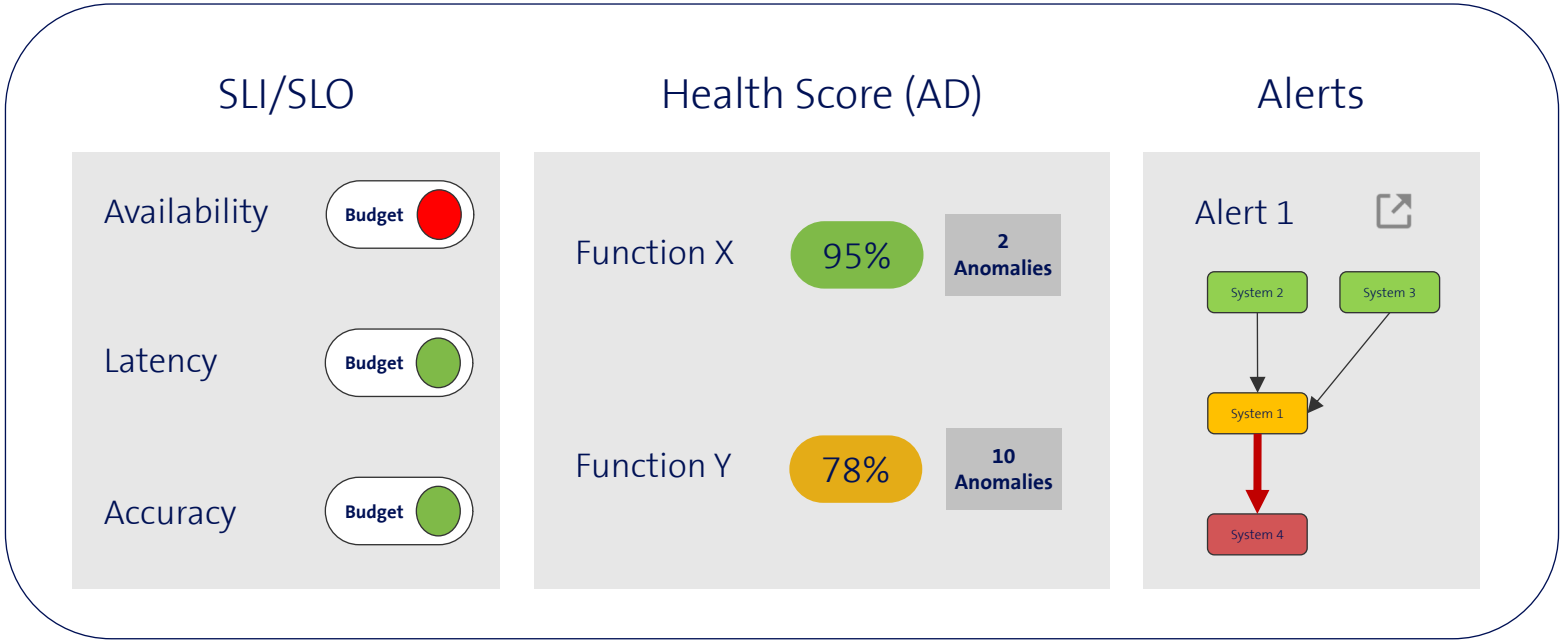
Conclusion

“The wheel is come full circle.”

-- King Lear

Answer: To AD or not to AD?

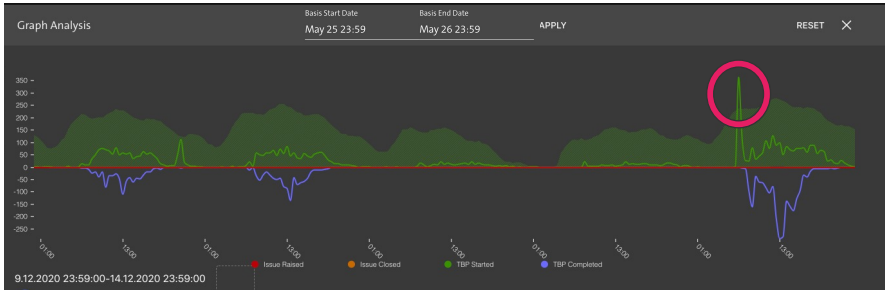
- Base your solution on the available data and resources ***and don't be scared to combine!***



Holistic Monitoring

- Goal is to understand the **problem chain end-to-end** and **from different perspectives**.
- Make dashboards with relevant information:
  - Problem **severity**;
  - **Data model** information;
  - Details from **external ticketing systems**.

Impact on Business and Customer Satisfaction



Real-time detection of problems with end-to-end monitoring

The End

Thank you for your attention!

Visit us during the **poster session** for more Q&A!

Joana Soares Machado

joana-soares-machado

Jelena Malić

jelena-malic