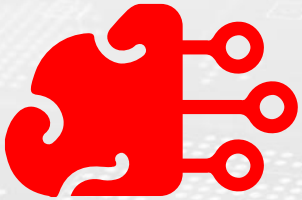
machinaide



Fuzzy time-series data modeling: theory and anomaly detection on Regatta® platform

Mechatronic Circus & Demo Day 2022

Riku Ala-Laurinaho
Postdoctoral researcher
riku.ala-laurinaho@aalto.fi

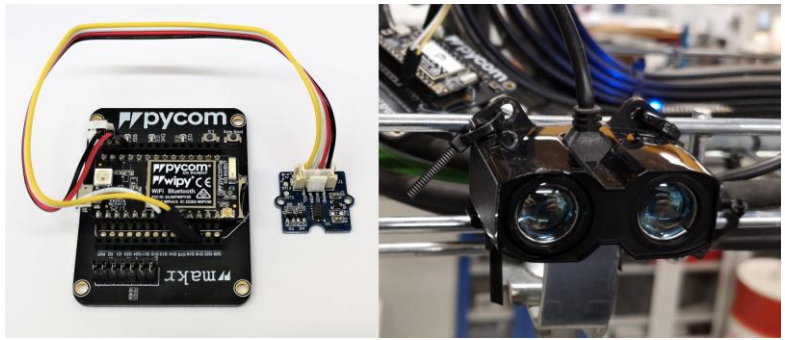
Aalto University

REMION

Miika Valtonen
CTO
miika.valtonen@remion.com

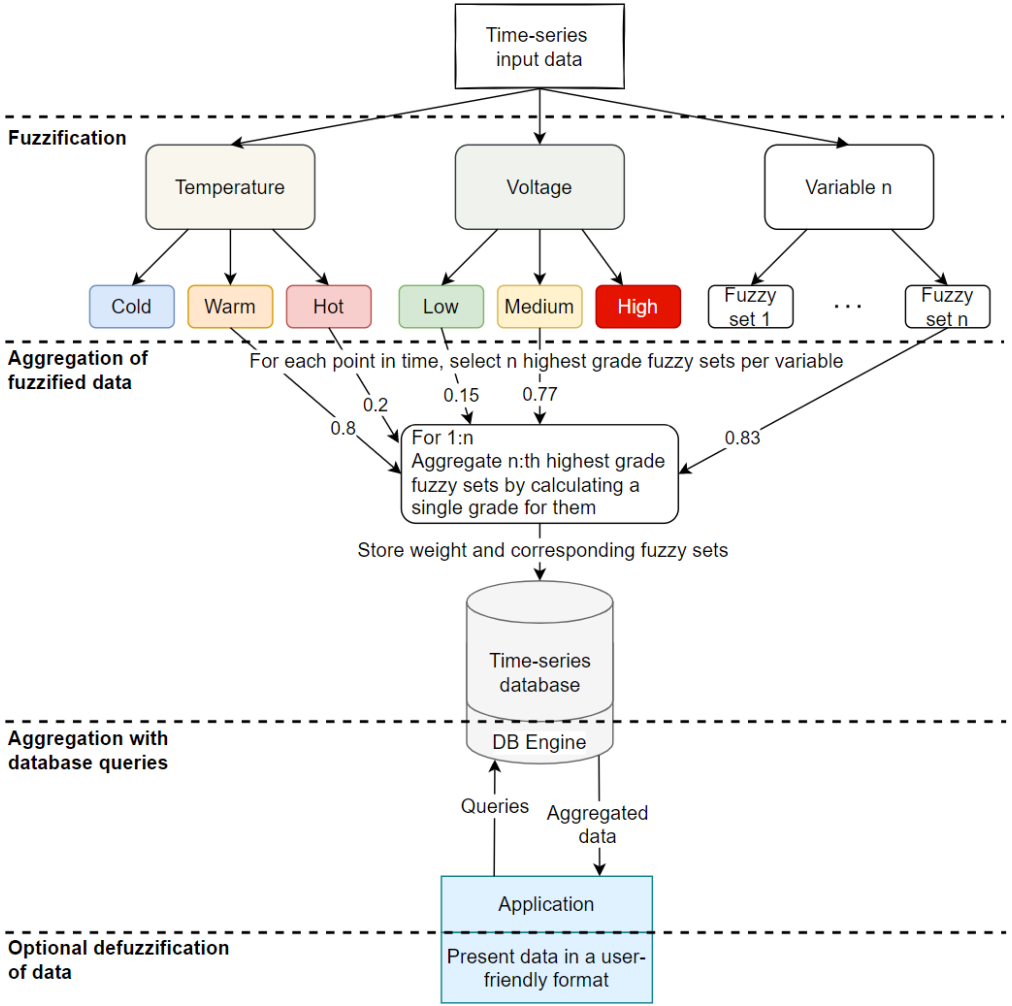
Background: modeling multivariate data

- Modern machines have tens or even hundreds of variables that are measured
- Problem: when **the number of variables increases, the computational complexity increases very quickly** (even exponentially), **problem-specific** models
- Solution: Generic **method to model machine data enabling diverse instant queries in time-signal space**
- Goal: provide information on how machines have been used and conditions that has led to failure



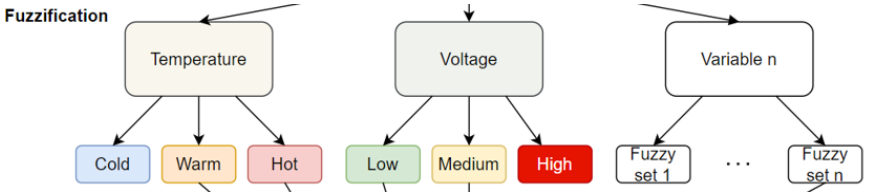
Ala-Laurinaho, R.; Autiosalo, J.; Tammi, K. Open Sensor Manager for IIoT. *J. Sens. Actuator Netw.* **2020**, *9*, 30. https://doi.org/10.3390/jsan9020030

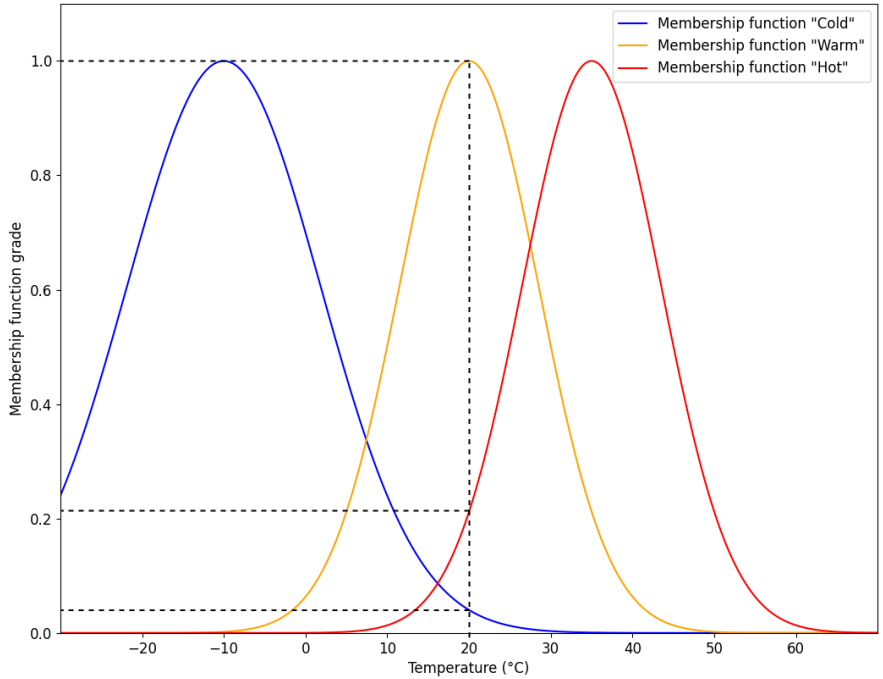
Overview of the data modeling method



Fuzzification

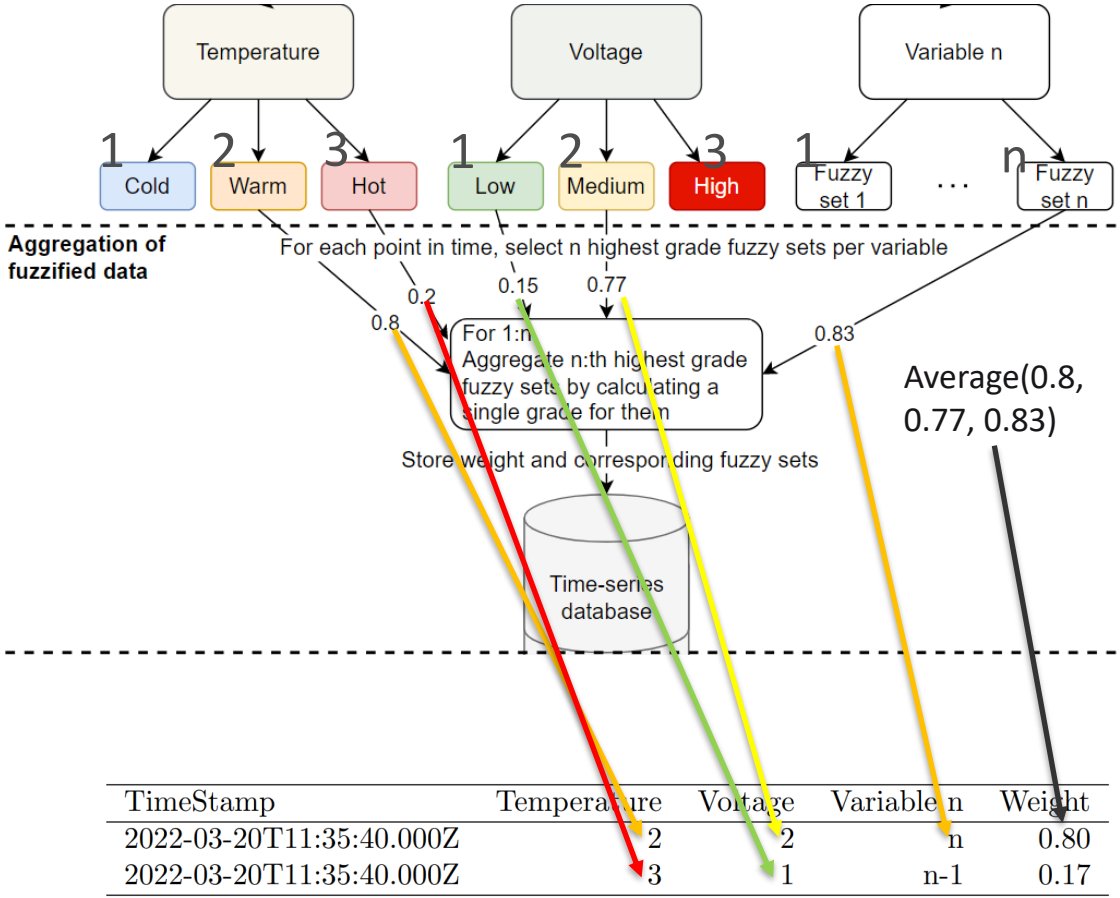
- **Fuzzy sets** are defined for each variable. For example, cold, warm, and hot for temperature
- Compared to Boolean logic, a value can belong to a fuzzy set with a **continuum of grades from 0 to 1**
- Number of fuzzy sets determines the accuracy of the model
- Using fuzzified variables for time-aggregation in later phases consumes **significantly less memory**





Aggregation to database

- For each point in time, **select** predefined number of the **highest grade fuzzy sets**
- **Combine the grades** of n:th highest grade of fuzzy sets as **weight** with different conjunction methods such as average, product, min, max, sum, etc.
- After spatial aggregation, data can be aggregated in **time-domain** by summing weights together in a **time bucket** -> more efficient queries



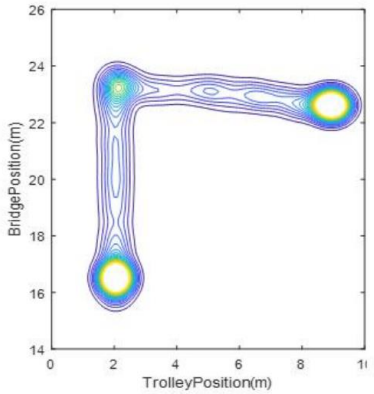
TimeStamp	Temperature	Voltage	Variable n	Weight
2022-03-20T11:35:40.000Z	2	2	n	0.80
2022-03-20T11:35:40.000Z	3	1	n-1	0.17

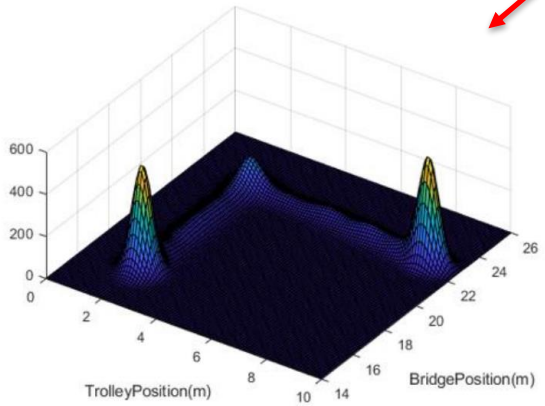
Database queries and visualization

- Queries allow **further aggregation of data**, for example, within certain time range
- Queried data can be used with **visualizations**, such as crane position

```
SELECT "TrolleyPosition", "BridgePosition", SUM(weight) AS sum_weight
FROM craneData
WHERE time BETWEEN '2020-01-30' AND '2022-03-20'
GROUP BY "TrolleyPosition", "BridgePosition"
ORDER BY sum_weight DESC
```

TrolleyPosition	BridgePosition	sum_weight
0	2	341.55
0	1	254.21
1	3	124.47
1	4	105.13
2	0	79.82
2	1	34.68





Overview | Data models | Concepts | Analysis

Analysis

Evaluate raw signal values against data model confidence scores

03.01.2022 00.00 - 07.01.2022 12.00

Data source

Model

Temperatures

Entity

Reach Stacker

Evaluated signal

Oil Temperature

on **temperature** dataset

Explaining signals

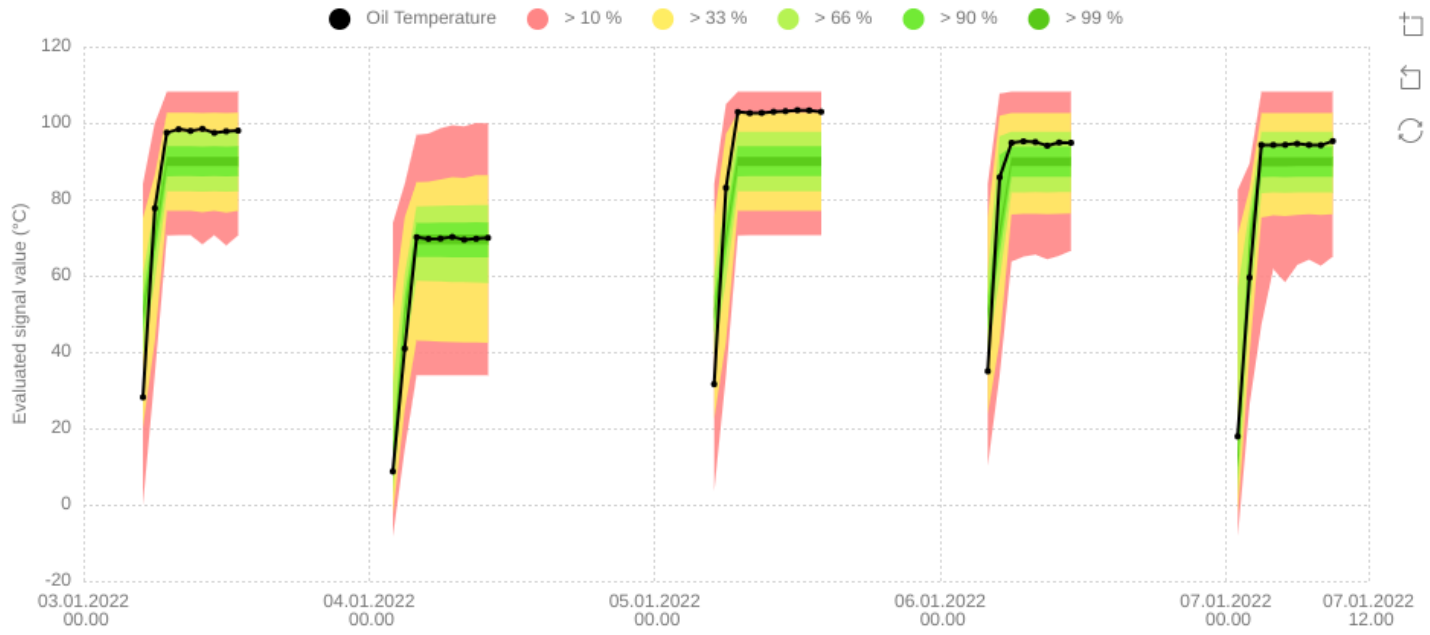
Cabin temperature

on **temperature** dataset

Engine Temperature

on **temperature** dataset

Time-series analysis | Entity distribution



Anomaly detection on Regatta® platform

1 Select data

- 10 Reach stacker machines, each operating ~8 hours per day in different climates
- Data from 1 year period, 7.4.2021 – 7.4.2022
- 3 temperature signals reported every 5 minutes

Reach stackers A-J

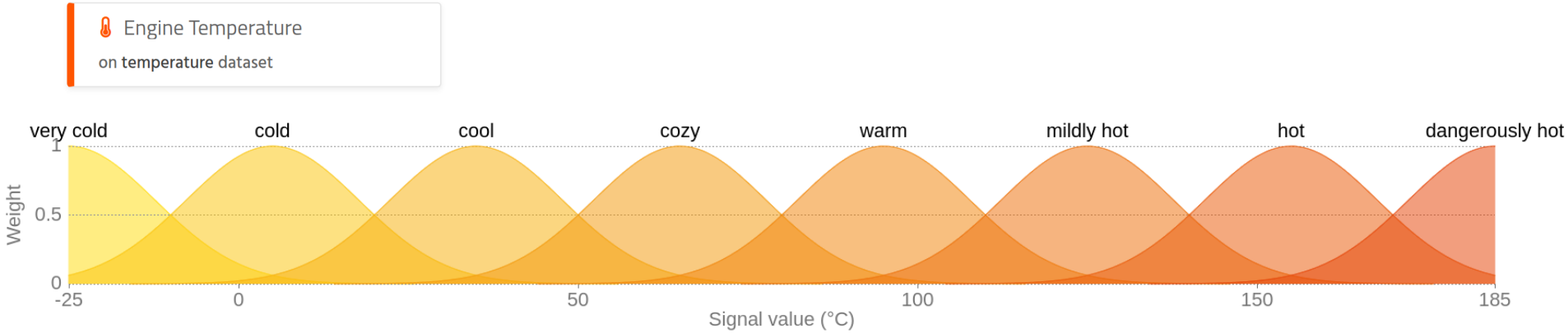
		Raw temperatures		
Time	Machine	Engine	Cabin	Oil
7.4.2021 19:05	A	26.1	27.2	26.4
7.4.2021 19:10	A	29.2	26.9	28.9
7.4.2021 19:15	A	26.2	26.8	36.3
7.4.2021 16:10	B	-0.8	0.7	-0.6
7.4.2021 16:15	B	6.4	1.8	5.5
7.4.2021 16:20	B	3.8	5.1	9.2

Raw data sample, total of 420k rows

❷ Define concepts

- Each signal is defined with Gaussian fuzzy terms
- Linguistic names help in data interpretation

Signal	Linguistic terms
Oil Temperature	very low · low · cold · normal · warm · high · extremely high
Cabin temperature	freezing · uncomfortable · cold · warm · hot
Engine Temperature	very cold · cold · cool · cozy · warm · mildly hot · hot · dangerously hot



3 Setup modeling parameters

Parameters

Model update and aggregation frequency *

7 days

Model aggregation function *

Mean value

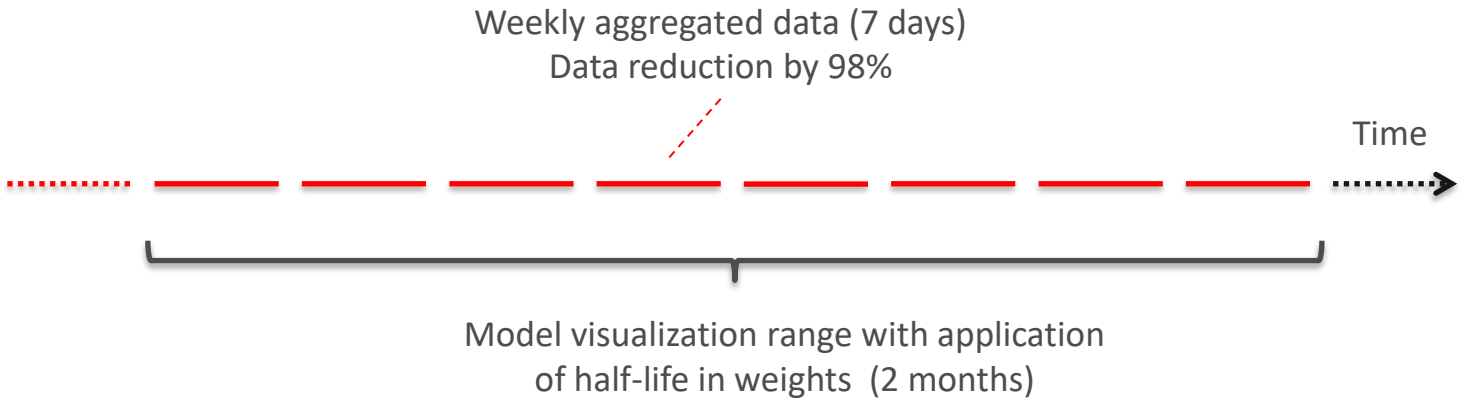
Data modeling interval *

2 months

What is the size of the **time bucket** and **model update frequency**?

What **time-aggregation function** should be used?

How old data is included in model queries?



4 Time-series analysis

Evaluated raw signal for machine A

Oil Temperature
on **temperature** dataset

Notes

- Machines have cooled down before start-up and **oil temperature warms up during the first hour of use**
- After that oil temperature maintains a **relatively level temperature**
- The operation of this machine shows **some anomaly during warmup but no anomaly during constant use**

Confidence score distribution

Confidence score ranges (0-100%) are predicted using other raw signals from the model at measurement time

Explaining signals

Cabin temperature
on **temperature** dataset

Engine Temperature
on **temperature** dataset

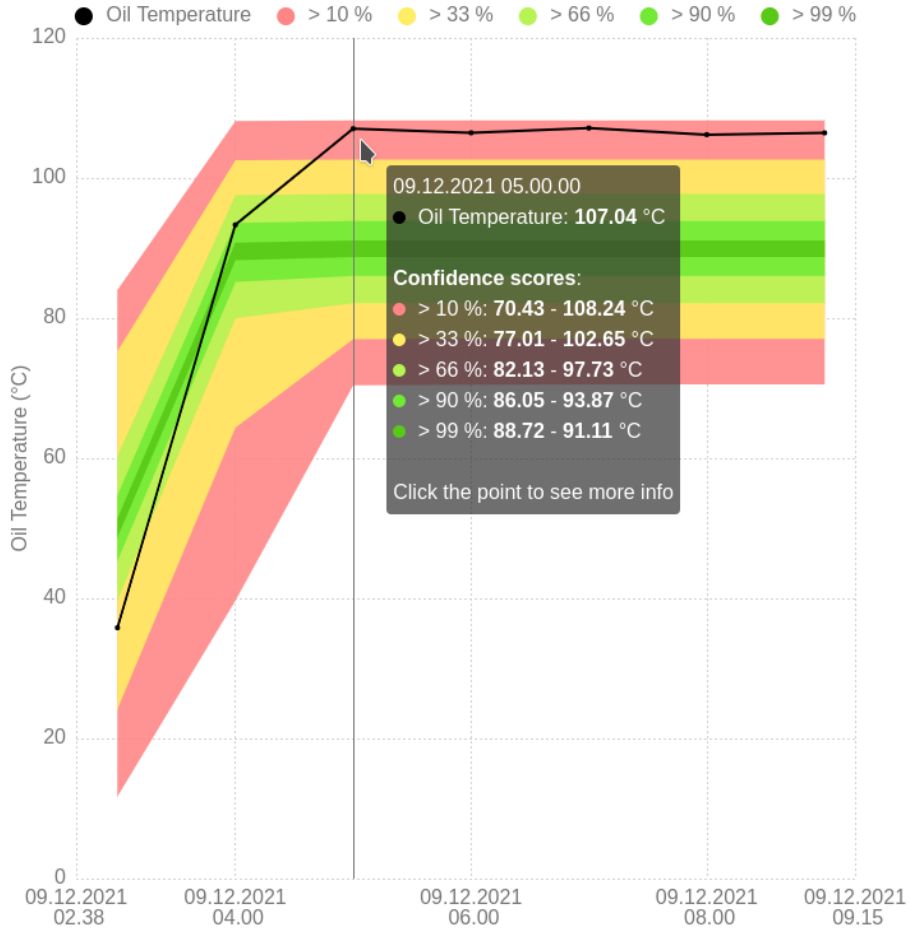
4 Time-series analysis

Notes

- As ambient air temperature changes, the **confidence score distribution rises**
- Machine shows **fairly normal operation during warmup, but oil temperature clearly overheats in constant use**

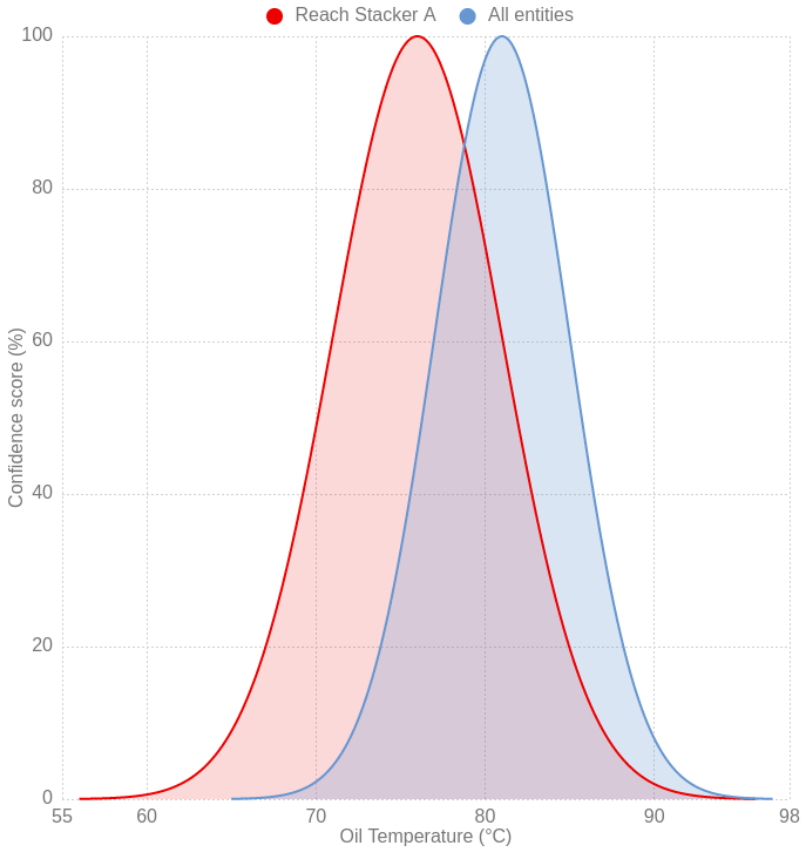
Time-series analysis view

- Allows its users to easily model the raw time-series data and **visually explore anomalies using confidence score distributions**



5 Model exploration

- As model data is updated periodically in the background, the model can be **visualized in realtime for predictive analytics purposes at fleet level**
- For example, **signal distributions from selected machines can be compared to other machines** operating at similar conditions, enabling early problem detection
- Model changes will enable **signal level predictions** and **proactive anomaly alerting** from all connected machines



Entity signal-distribution analysis view

Aalto University

12 500+
STUDENTS IN 2021

2 500+
INTERNATIONAL PEER-REVIEWED ARTICLES IN SCIENTIFIC JOURNALS IN 2021

REMION

Trusted Industrial IoT Ecosystem Partner

150+
IOT PROJECTS SINCE 2001

BILLIONS OF EUR
WORTH MACHINES AND DEVICES MONITORED