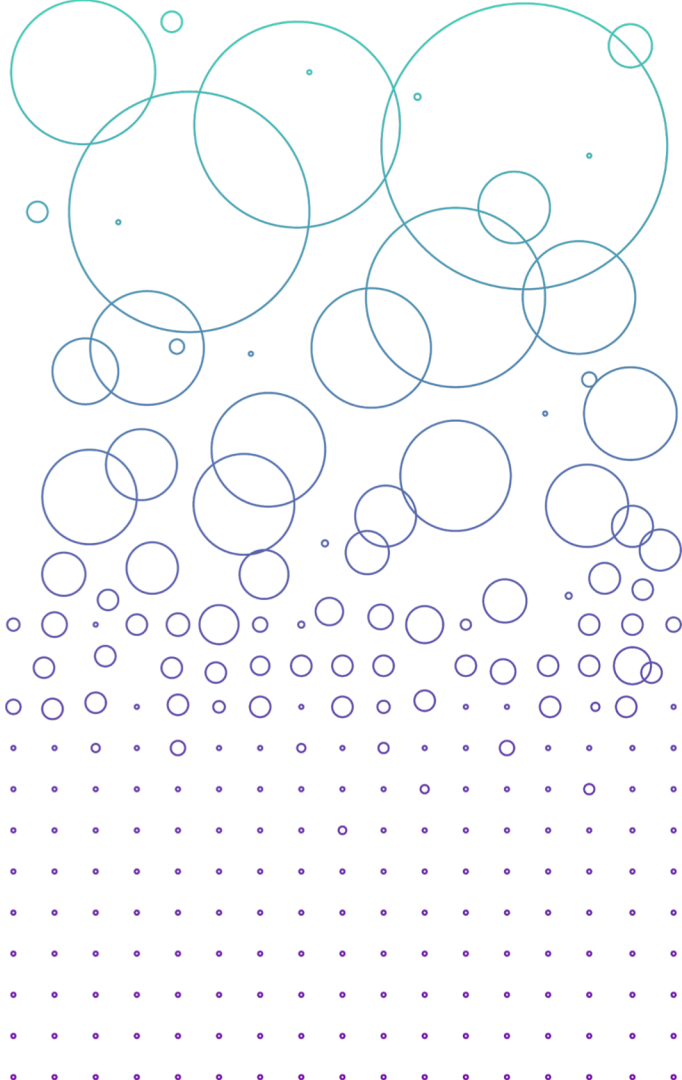
FLEET ANALYTICS

For Edge IoT solutions

Emmanuel Raj, Magnus Westerlund and Leonardo Espinosa-Leal

Emmanuel Raj |
TietoEvry |
emmanuelraj7@gmail.com
25/10/2020



My background

AI/ML

ARCADA

tieto EVRY

SILO AI

5+ years

Writing

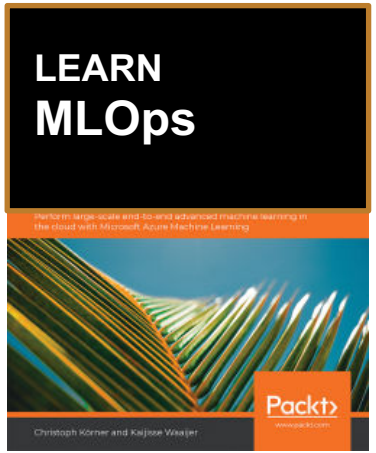
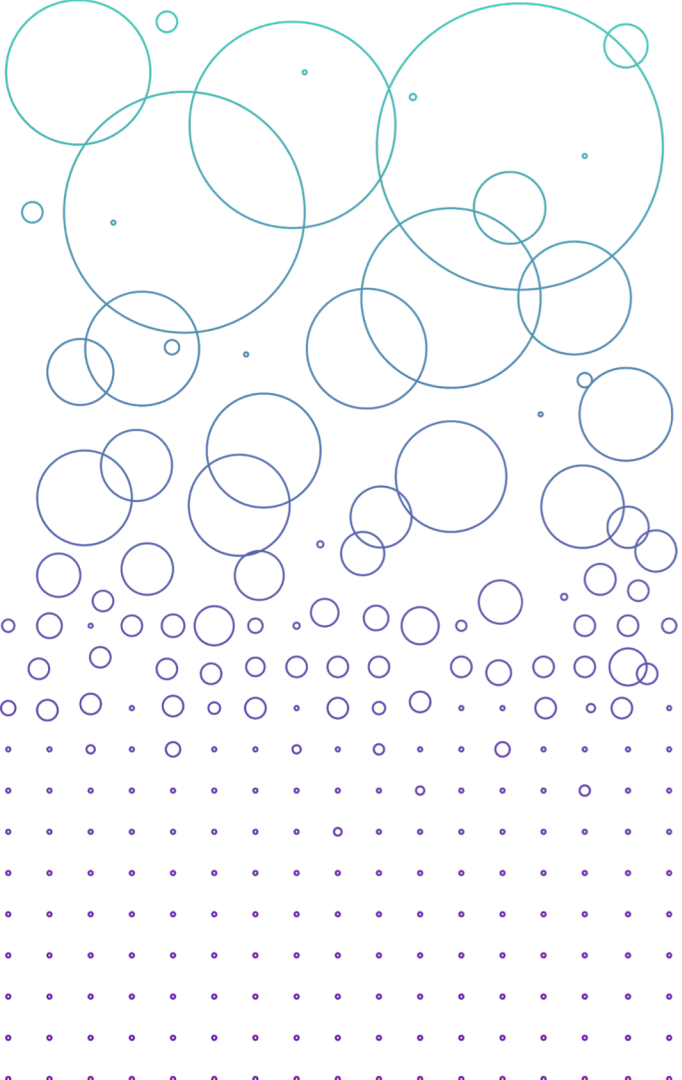


Table of Contents

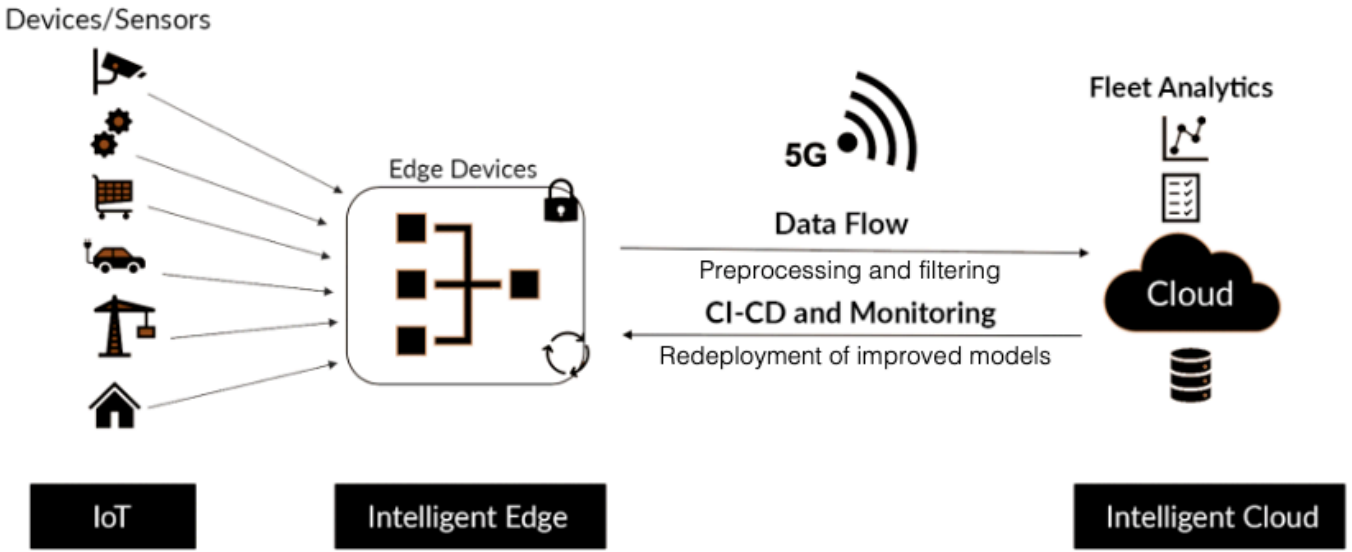
- Artificial Intelligence of things (AIoT)
- Fleet Analytics Framework for IoT solutions
- Experimental Framework Validation
- Reliability of fleet analytics framework

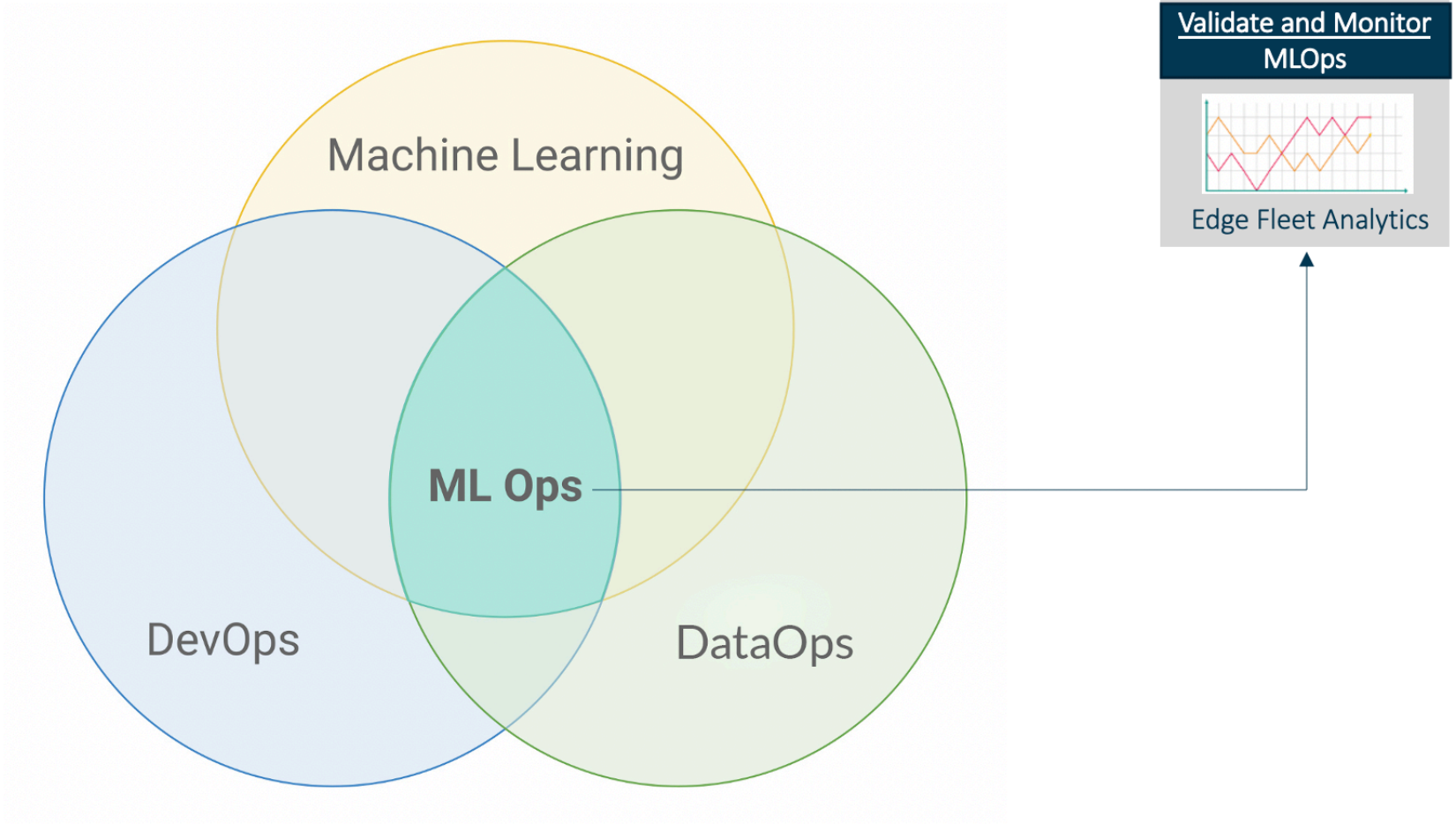
Slides posted on LinkedIn @emmanuelraj7



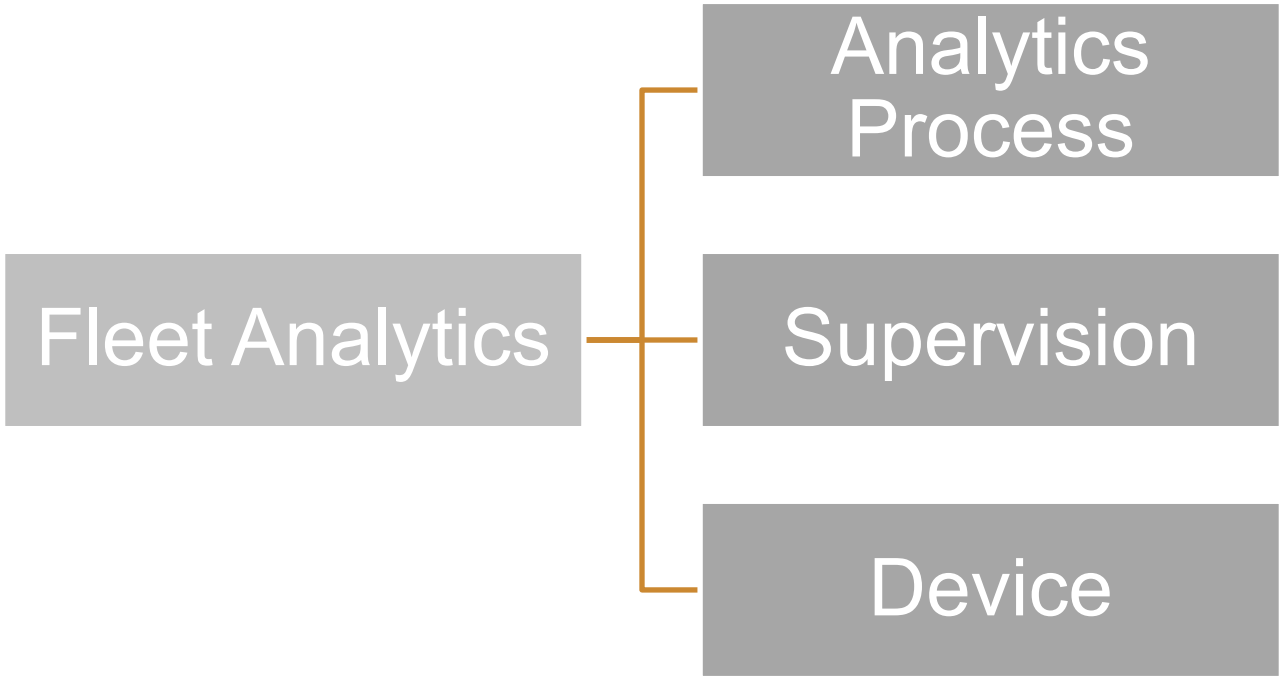
Artificial Intelligence of things (AIoT)

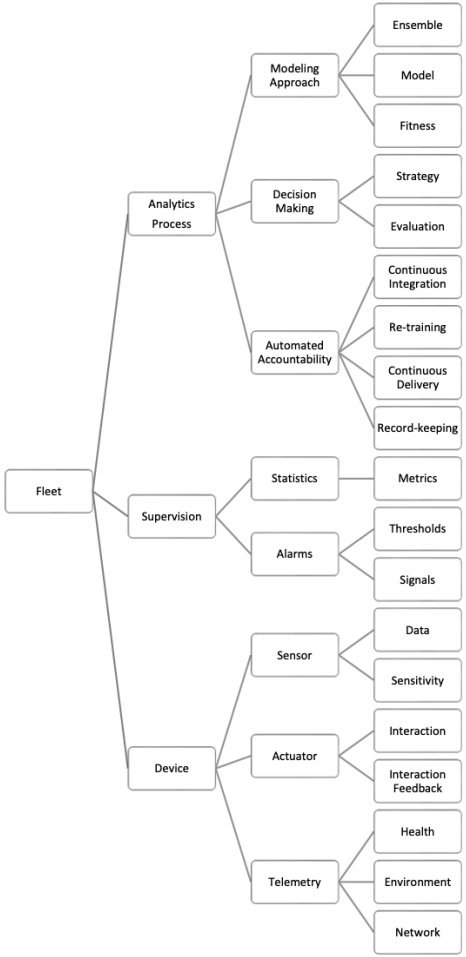
Artificial Intelligence of Things (AIoT) is the combination of artificial intelligence (AI) technologies with the Internet of Things(IoT) infrastructure to achieve more efficient IoT operations and decision making.

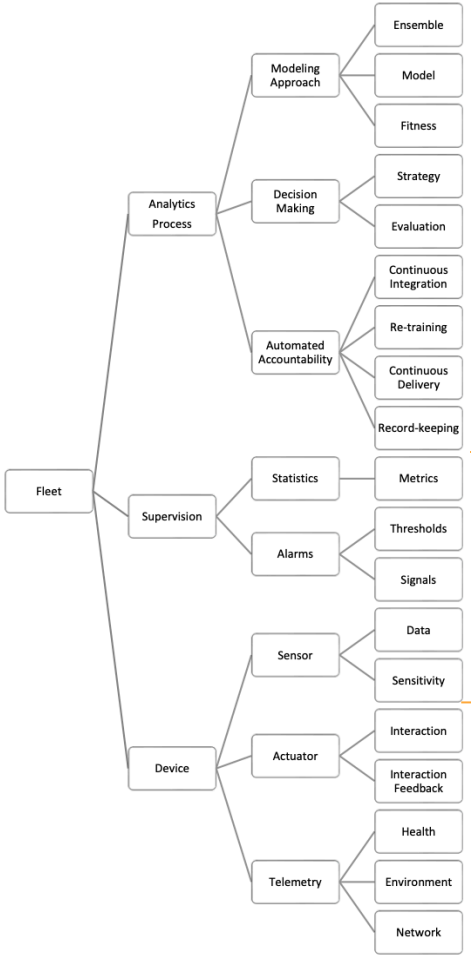


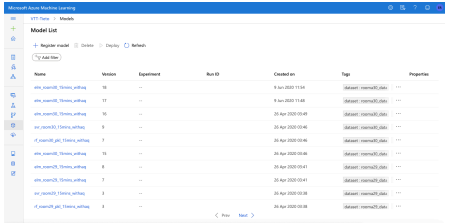

Validate and Monitor MLOps
Edge Fleet Analytics
Machine Learning
ML Ops
DevOps
DataOps

Fleet Analytics Framework

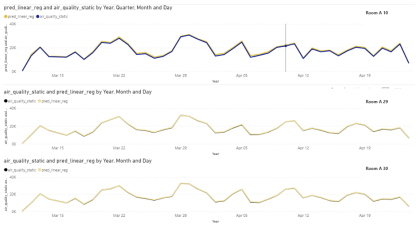

Fleet Analytics
Analytics Process
Supervision
Device


Fleet
Analytics Process
Modeling Approach
Ensemble
Model
Fitness
Decision Making
Strategy
Evaluation
Automated Accountability
Continuous Integration
Re-training
Continuous Delivery
Record-keeping
Supervision
Statistics
Metrics
Alarms
Thresholds
Signals
Device
Sensor
Data
Sensitivity
Actuator
Interaction
Interaction Feedback
Telemetry
Health
Environment
Network

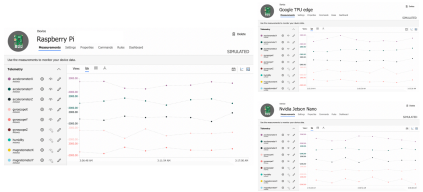




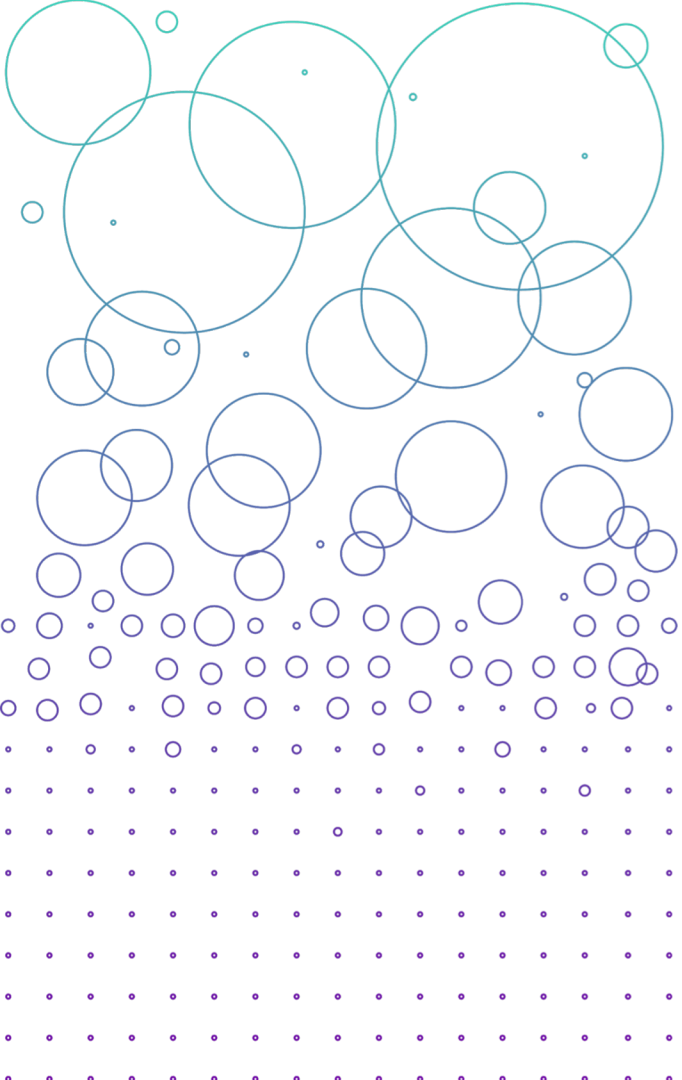
Analytics process



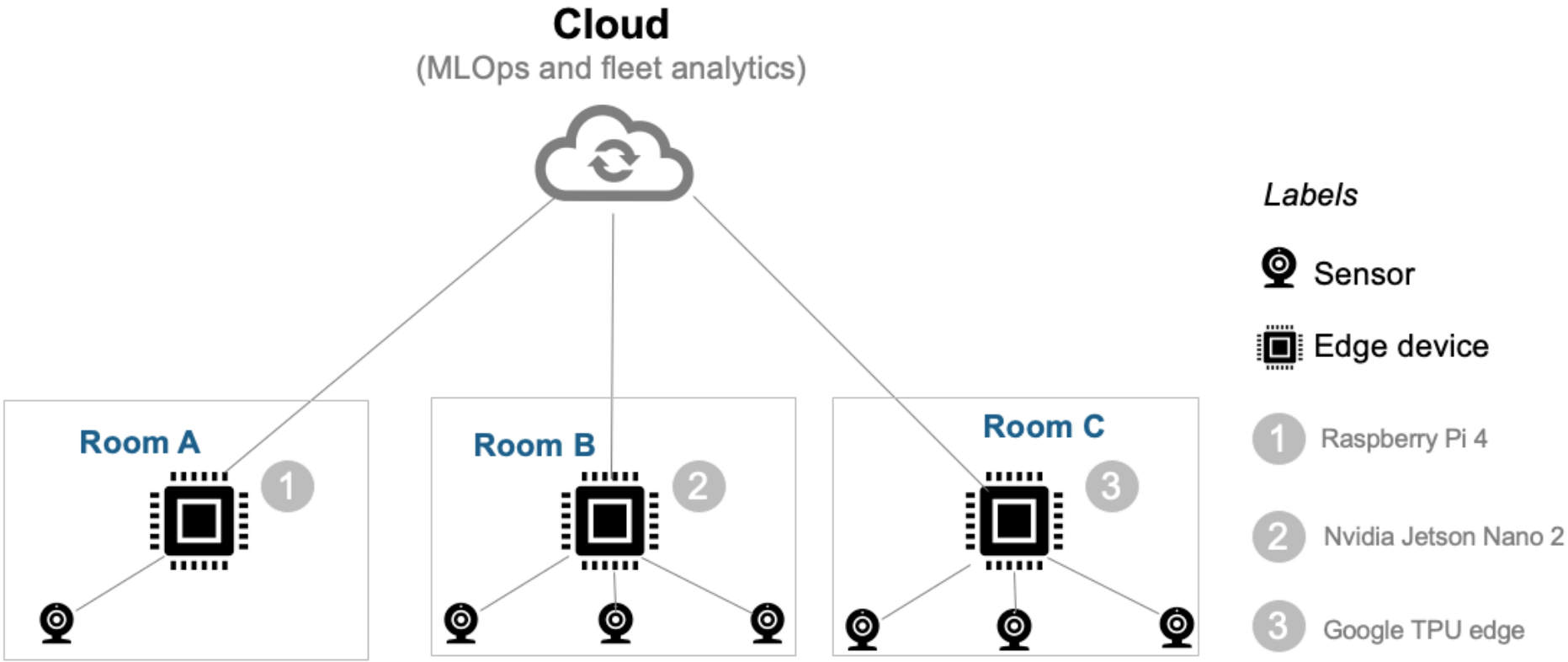
System supervision



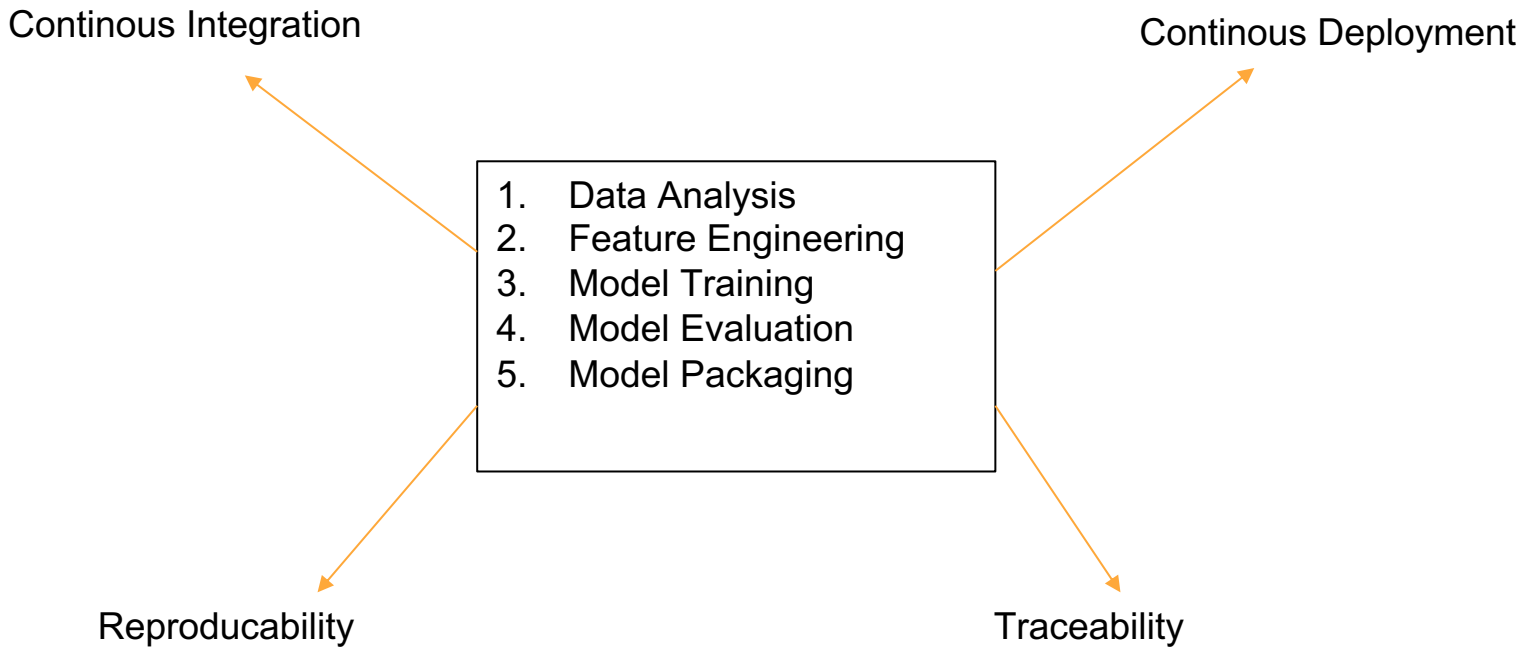
Device Analytics



Experimental Framework Validation


Cloud
(MLOps and fleet analytics)
Room A
1
Room B
2
Room C
3
Labels
Sensor
Edge device
1 Raspberry Pi 4
2 Nvidia Jetson Nano 2
3 Google TPU edge

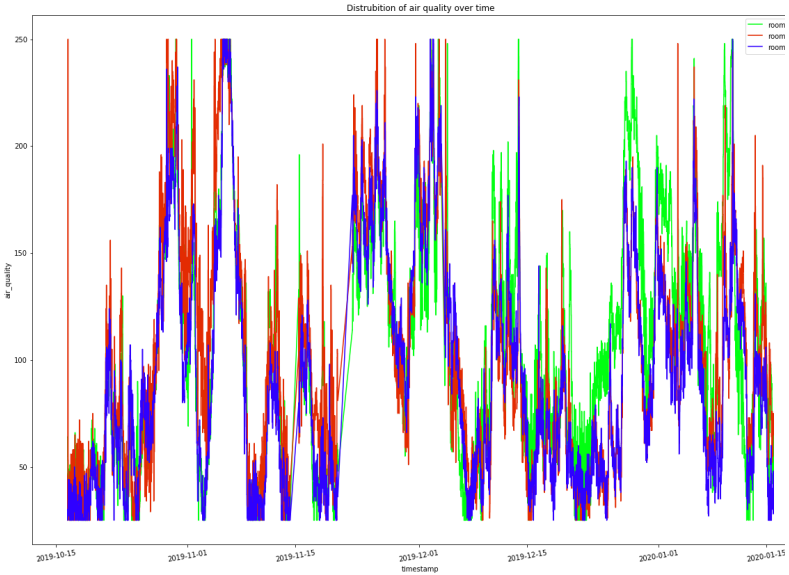
MLOps for AIoT

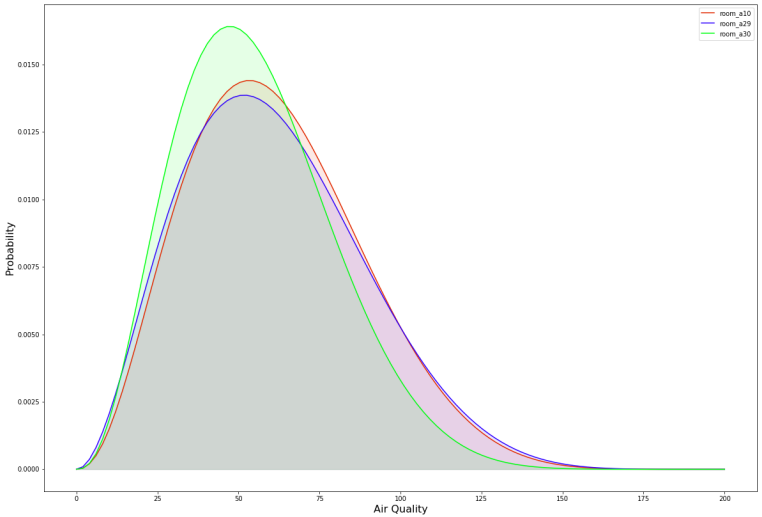


Data Analysis

timestamp	name	room	room_type	floor	air_quality	air_quality_static	ambient_light	humidity	iaq_accuracy	iaq_accuracy_static	pressure	temperature
2019-10-16 11:45:12	T009	A09	office_room	A	31.0	27.0	10.0	34.32	1.0	1.0	1010.0	21.57
2019-10-16 11:45:12	T010	A10	office_room	A	64.0	42.0	82.0	33.93	1.0	1.0	1010.0	22.95
2019-10-16 11:45:28	T017	A17	office_room	A	25.0	25.0	123.0	34.71	1.0	1.0	1010.0	22.52
2019-10-16 11:45:40	T019	A19	office_room	A	51.0	40.0	1.0	34.71	3.0	3.0	1009.0	22.00
2019-10-16 11:45:45	T020	A20	office_room	A	59.0	39.0	6.0	33.15	1.0	1.0	1010.0	22.04

Timeline - 3 months (15-10-2019 to 15-01-2019)
Total 537873 number of rows or events were recorded.
Size of the data: 45.9 MB


Distrubition of air quality over time
room
room
room
air_quality
timestamp
2019-10-15
2019-11-01
2019-11-15
2019-12-01
2019-12-15
2020-01-01
2020-01-15


room_a10
room_a29
room_a30
Probability
Air Quality

Model Training & Evaluation & Packaging

Multivariate Time Series Forecasting

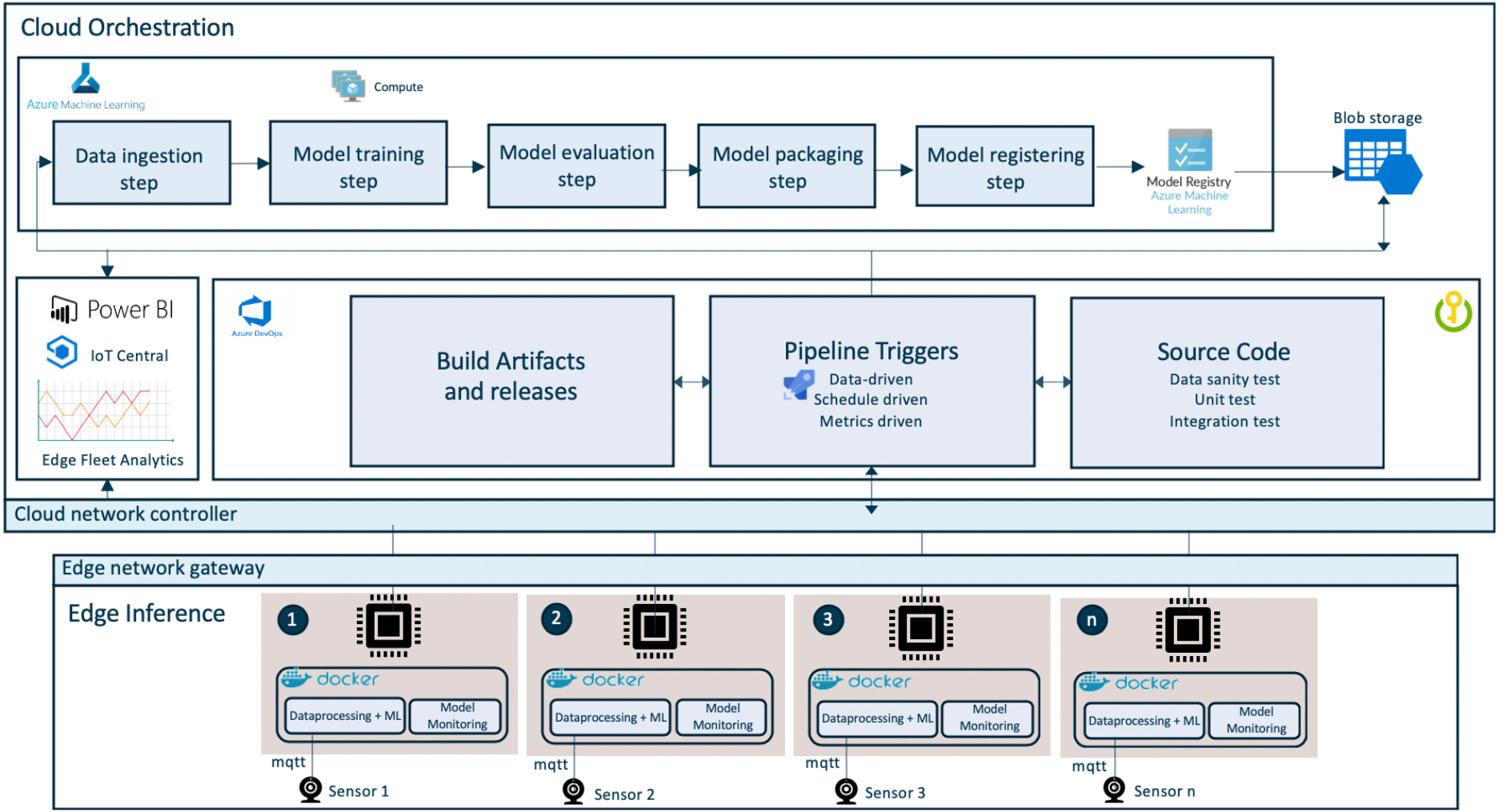
1. Linear regression
2. Extreme learning machines
3. Random forest regressor
4. Support vector regressor

Model Training Results			
Room name	Algorithm	Cross Validation RMSE (train)	Test RMSE
Room A10	MLR	5.020	5.875
Room A10	ELM	6.325	6.208
Room A10	RFR	10.710	9.987
Room A10	SVR	6.046	5.977
Room A29	MLR	5.362	4.158
Room A29	ELM	11.202	4.223
Room A29	RFR	11.676	9.208
Room A29	SVR	8.073	4.176
Room A30	MLR	3.648	3.551
Room A30	ELM	7.920	3.895
Room A30	RFR	9.686	7.720
Room A30	SVR	5.177	3.55

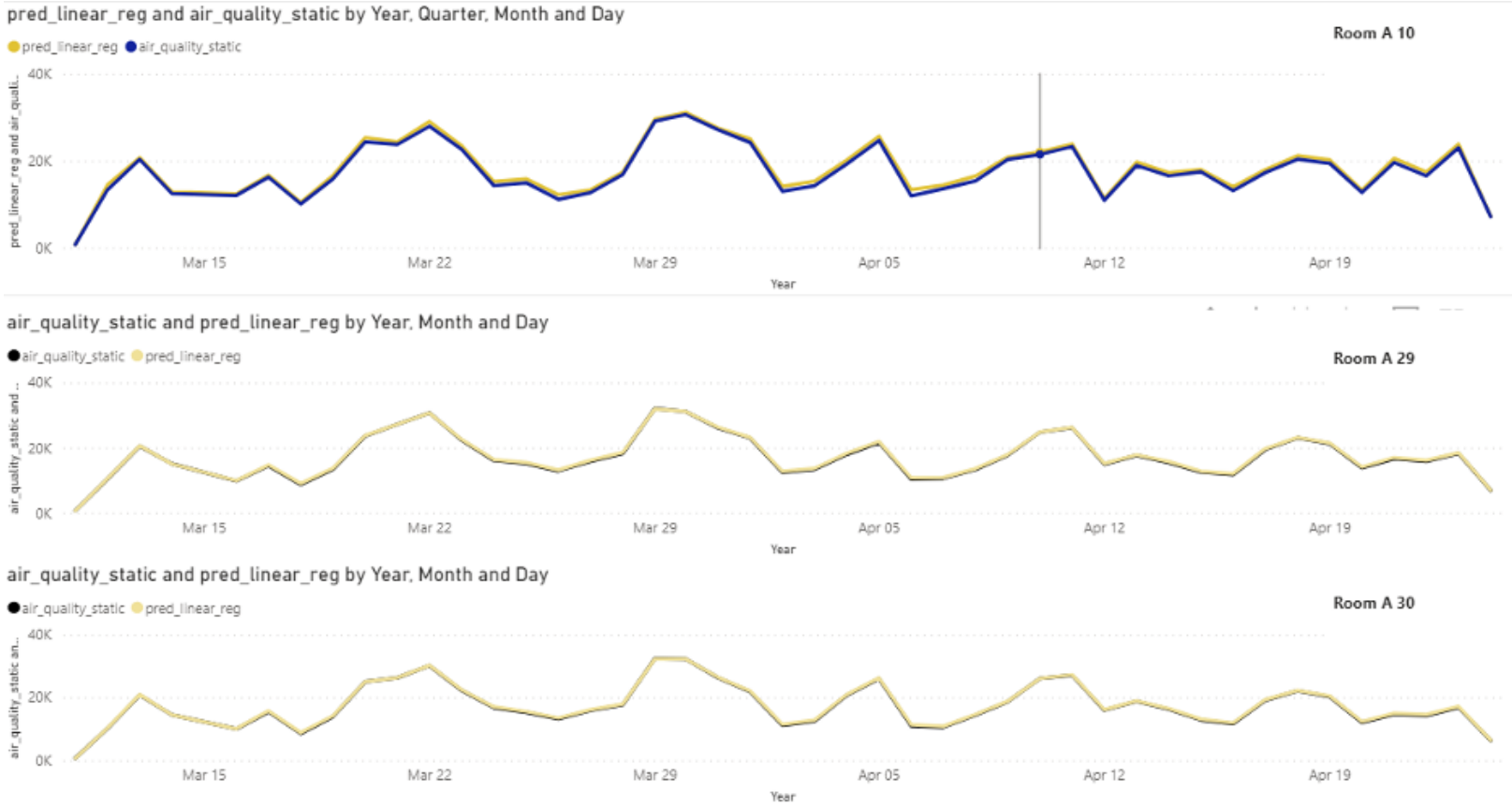
Table 2. Model training results.

ONNX

Reliability of Fleet Analytics framework


Cloud Orchestration
Azure Machine Learning
Compute
Data ingestion step
Model training step
Model evaluation step
Model packaging step
Model registering step
Model Registry
Azure Machine Learning
Blob storage
Power BI
IoT Central
Edge Fleet Analytics
Azure DevOps
Build Artifacts and releases
Pipeline Triggers
Data-driven
Schedule driven
Metrics driven
Source Code
Data sanity test
Unit test
Integration test
Cloud network controller
Edge network gateway
Edge Inference
1
2
3
n
docker
Dataprocessing + ML
Model Monitoring
mqtt
Sensor 1
Sensor 2
Sensor 3
Sensor n

Edge inference



Analytics process

- 23 new model deployed.
- For room a10-raspberry pi had 7 new ML models deployed, room a29-jetson nano had 7 models deployed and room a30-google TPU had 9 models deployed during the experiment timeline.

Supervision

- In case of future air quality forecasted above 100 aqi the system would alert the users (building maintenance personnel) to regulated air quality in the rooms

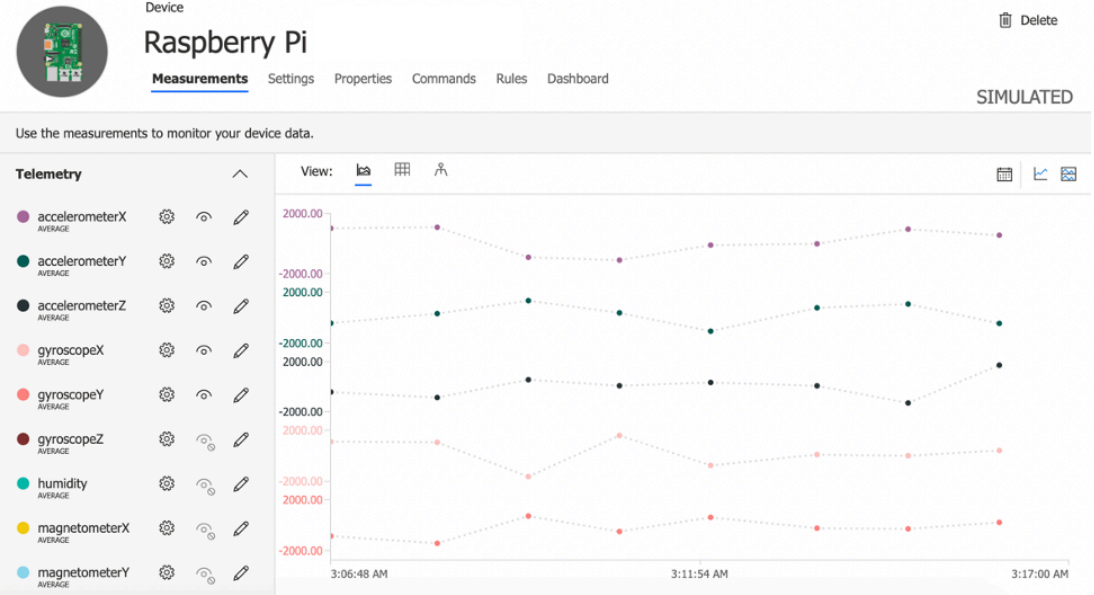
Device

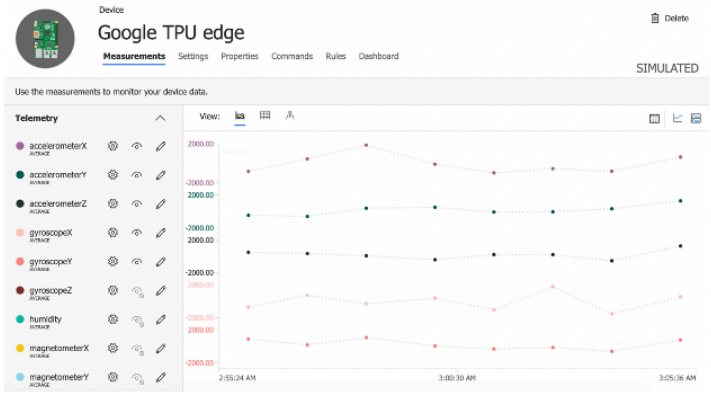
- Device analytics provided an overview of device performance over some time with telemetry data like accelerometer, gyroscope, humidity, magnetometer, pressure, and temperature.
- All edge devices' performance was stable without any device failures.
- Costs, inference speed and latency.

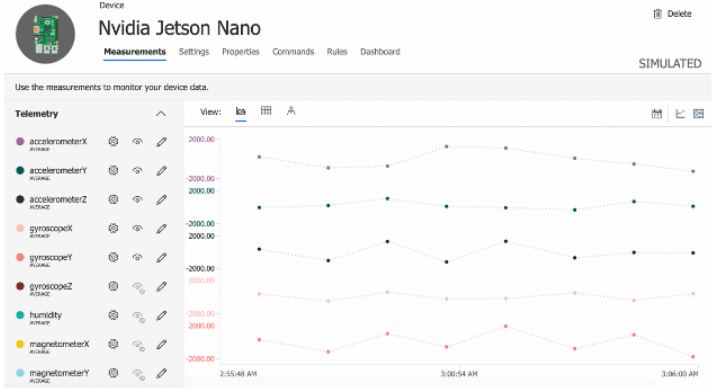
Realtime machine learning inference at the edge					
S.no	Date of model change	Edge Device	Deployed Model	Model Drift (RMSE)	Model Retrain (RMSE)
1	15-03-2020	Jetson nano 2	ELM	16.39	4.1
2	16-03-2020	Google TPU edge	RFR	14.23	6.3
3	16-03-2020	Raspberry pi 4	MLR	11.91	4.3
4	17-03-2020	Raspberry pi 4	ELM	13.27	8.1
5	22-03-2020	Jetson nano 2	SVR	22.32	6.2
6	24-03-2020	Google TPU edge	RFR	17.11	4.4
7	27-03-2020	Raspberry pi 4	MLR	16.22	4.7
8	29-03-2020	Jetson nano 2	ELM	30.28	8.2
9	30-03-2020	Google TPU edge	SVR	18.12	5.4
10	05-04-2020	Raspberry pi 4	MLR	12.92	3.2
11	10-04-2020	Jetson nano 2	SVR	17.21	5.2
12	11-04-2020	Google TPU edge	MLR	13.42	4.7
13	13-04-2020	Jetson nano 2	ELM	27.29	5.3
14	17-04-2020	Google TPU edge	RFR	17.46	6.9
15	19-04-2020	Raspberry pi 4	SVR	16.32	5.1
16	19-04-2020	Google TPU edge	MLR	11.91	3.4
17	21-04-2020	Jetson nano 2	ELM	23.26	7.3
18	22-04-2020	Google TPU edge	RFR	16.92	7.2
19	24-04-2020	Raspberry pi 4	SVR	17.87	5.2
20	25-04-2020	Google TPU edge	MLR	13.92	5.2
21	25-04-2020	Jetson nano 2	SVR	19.21	7.9
22	26-04-2020	Raspberry pi 4	ELM	23.57	6.4
23	26-04-2020	Google TPU edge	SVR	18.21	5.5

Table 3. AIoT experiment machine learning inference results.

Device Analytics







Edge Inference vs Cloud Inference – Experiment Results

We did a research experiment was done for a month to predict room air quality for 10 rooms, each room had 1 edge device for inference. Simultaneously we tested with cloud inference.

	Edge Devices (10)	Cloud (1 Nodes)
Device	Raspberry Pi 4	**DS2_v2** (Azure)
Computation	40 vCPUs (4 x 10)	2 vCPUs
RAM	40GB (4 x 10)	7 Gb
Temporary Storage	640 GB (64GB/device)	14 GB
Data Pruned	22%	0%
ML inference/minute	1/device	10
Average inference time	**~0.2 sec**	**~2.2 Sec**
Total Cost/Month	**~ € 10/month**	**~ € Euro 93/month**

Costs calculations as per: https://www.raspberrypi.org/forums/viewtopic.php?t=18043, https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/

Cost reduction

9x less

Inference Speed

9x Faster

Data Pruned

22%

Cloud Latency

2 Seconds

Edge Inference vs Cloud Inference - Scaled

Research experiment was done for a **month** to predict room air quality for 10 rooms, each room had 1 edge device for inference. Simultaneously was tested with cloud for inference.

	Edge Devices (1000)	Cloud (100 Nodes)
Device	Raspberry Pi 4	**DS2_v2** (Azure)
Computation	4000 vCPUs (4 x 1000)	200 vCPUs (2x100 nodes)
RAM	4TB (4 x 1000)	700 Gb (7 x 100 nodes)
Temporary Storage	64TB (64GB/device)	1.4TB (14 x 100 nodes)
Data Pruned	22%	0%
ML inference/minute	1	1000
Average inference time	**~0.2 sec**	**~2.8 Sec**
Total Cost/Month	**~ € 1000/month**	**~ € 9300/month**

Costs calculations as per: https://www.raspberrypi.org/forums/viewtopic.php?t=18043, https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/

Thank you!

emmanuelraj7@gmail.com