MOdel based coNtrol framework for Site-wide OptimizatiON of data-intensive processes

D7.3 - Initial Demonstrator in the Aluminium and Plastics Domains

Deliverable ID	**D7.3**
Deliverable Title	**Initial Demonstrator in the Aluminium and Plastics Domains**
Work Package	**WP7 – Demonstration and Evaluation in the Aluminium and Plastics Domains**
Dissemination Level	**PUBLIC**
Version	**1.3**
Date	**2019-02-28**
Status	**Final**
Lead Editor	**AP**
Main Contributors	**A. Bernard, A. Auge, V. Maigron (AP)** **M. De Pieri (LCEN)** **V. Bonnivard (PROBAYES)** **N. Kolokas, A. Pasias (CERTH)** **R. Schlutter (KIMW)** **M. Dias (GLN)** **S. Devasya (FIT)**

Published by the MONSOON Consortium

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723650.

Document History

Version	Date	Author(s)	Description
0.0	2018-05-28	A. Bernard (AP)	First Draft with TOC
0.1	2018-06-19	A. Bernard (AP)	Introduction + First draft of the paragraph for the definition of Proof of Concept for the aluminium domain
0.13	2018-12-13	M. De Pieri (LCEN)	Contribution to section 2.1 and 4.1.2
0.2	2018-12-16	V. Bonnivard (PROB)	Contribution of PROB
0.22	2018-12-17	N. Kolokas (CERTH) A. Pasias (CERTH)	Contributions to subsections 4.1.1.1, 4.1.1.3
0.4	2018-12-22	R. Schlutter (KIMW) M. Dias (GLN)	Contribution of GLN & KIMW
0.5	2019-01-13	A. Auge (AP) A. Bernard (AP)	AP's modification after internal reviews: formatting, minor corrections
0.99	2019-02-05	V. Maigron	Integration of all partners contribution Contribution for AP Pre review version
1.0	2019-02-22	V. Maigron	Modification after KIMW review
1.1	2019-02-25	S. Devasya (FIT) M. Dias (GLN)	Plastic Domain inputs: chapter 3 review and chapter 4 contribution
1.2	2019-02-26	V. Maigron	Version ready for internal review
1.3	2019-02-28	V. Maigron	Included feedback from reviewers Final version submitted to the EC

Internal Review History

Version	Review Date	Reviewed by	Summary of comments
1.2	2019-02-26	M. Dias (GLN)	Accepted with minor layout and grammatical corrections
1.2	2019-02-28	R. Schlutter (KIMW)	Accepted with minor layout and grammatical corrections

Table of Contents

1 Introduction

This deliverable documents the process followed to develop the use cases in the Aluminium and Plastics domains. The development of use cases consists of 5 steps:

- The Proof of Concept (POC) which aims to test high-level business case assumptions in real conditions. The use cases will be detailed within framed project charters, including the main data to leverage and the KPIs foreseen for the evaluation. At the end of this first step, the performances of the initial model developed (POC results) will be able to be assessed.
- Proof of Value (POV)per use case in order to pilot and verify the business case assumptions (€, HSE) before launching the next steps
- Proto or Feasibility per use case, which corresponds to thebuilding of the Proto solution. This phase will include the development of real time connectors to retrieve data, the first tests and simulations in R&D context, the packaging of the developed predictive function and its simulation, performance test and validation
- Demo or Deploying for real time operation on the corresponding pilot site. At this stage of the use case, the deployment of the predictive function in the runtime container might be able. During this phase, the handover to the operational teams should also be performed.
- Deploying the model in different environments. This step is not in the scope of the MONSOON project.

The last two steps are not part of the task 7.2. The Demo phase is linked to the task 7.2for aluminium and task 7.3 for plastic. The Deploy phase is not in the scope of the MONSOON project.

Because it is the first version of the document, it focuses on the activities performed and results obtained during the first iteration of the MONSOON Project.

1.1 Scope

In a first chapter paragraph, this deliverable will detail the POC that was defined for each use case and each domain, describing also the process followed to determine the engagement charters that were the basis for a common understanding of the scenario to be studied. Each POC will be also introduced in these first paragraphs.

The second chapter will be more oriented to POV phase. It will illustrate how the impacts of the models or predictive functions that were developed during this first iteration are assessed.

The third chapter will focus on the exploitation of the results, explaining the activities led during the first iteration and giving the performances of the algorithms developed so far.

The last chapter of this deliverable will tackle the training aspects, depicting the organization and the resources to be implemented for this activity.

1.2 Related documents

ID	Title	Reference	Version	Date
[RD.1]	Initial Semantic framework for dynamic multi-scale industry modelling	D4.1	1.0	2017-11-28
[RD.2]	Initial Life Cycle Management plugin	D5.7	1.2	2017-12-22
[RD.3]	Initial Evaluation Framework	D7.1	1.2	2018-03-28
[RD.4]	Process Industry Domain Analysis and Use Cases	D2.2	1.3	2016-12-30

2 Definition of the Proof of Concept (POC)

In order to describe the Proof of Concept, the development of a structured document was started -the engagement charter. It was first shared amongst the industrial partners, but it aims to be used as a common basis for the discussions between MONSOON partners (industrial, data scientist, IT…). This document clarifies the expected use case, also called afterwards scenario. One document for each use case will be set forth. Eachengagement charter contains a first description of each scenario, as detailed as possible given when it is written (at the early stage of the study). This tool will be completed by the semantic model performed on the semantic platform (see D4.1 for more information regarding the semantic framework).

The common structure of the document is described below:

- The first part called "Problem statement" gives the context of the scenario to be studied (the impacts on the production processes and how it is currently handled, in particular).
- The second part of the document, "Aim – Desired end state" specifies the objectives from the industrial or end-user point of view. The industrial partner can provide the different steps for studying the scenario based on his experience. This might be enhanced by the information contained in the fifth part "Desired outcomes / Success criteria"
- To complete this second part, the industrial partner will be able to detail the "Benefits" of the scenario (OPEX/CAPEX, HSE, Quality…)
- Afterwards, the partners must clarify what is in the scope of the defined study and what is out of the scope (Data sources, equipment, operations, documentation…)
- Like for each study or development, the performance of the outcomes must be assessed, in our case the models or algorithms. For this purpose, the industrial partner must determine KPIsin "Measurement" section.
- The next parts are linked to the needed human Resources and the estimated Workloadfor the study. The objective of this section is to be able to verify the compliance with the estimated workload given in the Grant Agreement, and eventually address resource issues.
- The engagement charter provides also an estimated planning for the development of the use case.
- Finally, a risk analysis is included in this document presenting the preventive actions to manage the risks.

In the coming sections, some examples of scenarios and related Engagement charters are described.

2.1 Aluminium domain

2.1.1 Description of the process to obtain the Engagement charters

In the aluminum domain, a first scenario was chosen and depicted with the help of the domain director who has a broader view of the Client's needs. The draft is then discussed with the pilot site representatives to confirm the detailed use case corresponds to one of their current problem, or to align the use case if necessary.If the pilot site representatives are not available, the process expert can somehow replace them to define the problem statement (based on the knowledge of the pilot processes and abnormal behaviour).

Afterwards, it is an iterative process with the Consortium members that will be involved in the use case study, in order to clarify the unclear points or elements of the engagement charter and estimate the workload taking into the initial description of the use case.

In the aluminium domain, 4 different engagement charters were defined, 3 in the Carbon area and 1 in the Electrolysis area:

- Carbon area
    - Equipment stoppages (POC1)
    - Equipment misbehaviours (POC2)

    - Prediction of the anode quality (POC3)
- Electrolysis area (POC4) – 3 sub-scenarios
    - Prediction of the liquid heights
    - Prediction of the pot thermal balance
    - Prediction of the anode effects

In the next sub-section, each scenario will be introduced and taken as a basis the corresponding engagement charter.

2.1.2 Description of the scenarios and explanation regarding the planning

2.1.2.1 Equipment stoppages (POC1)

2.1.2.1.1 Problem Statement

Anodes contribute as the biggest variable impact to the aluminium production cost.

Optimization goals may focus on the anode paste plant to produce better anodes in terms of quality. The optimized anode characteristics are listed hereunder:

- Higher consistency of anode properties (homogeneous anode batches, in particular regarding the chemical properties, the mechanical properties are amongst the following characteristics)
- No anode cracking or slabbing
- No mushrooms in pot
- Good anode current distribution
- Low anode voltage drop
- Low consumption figures
- High current efficiency

These efforts take part of an overall goal for the carbon area which aims to:

- Increase throughput
- Reduce scrap
- Reduce energy consumption
- Reduce breakdowns and maintenance costs

The recent technical reports concluded that Aluminium Dunkerque plant produces anodes of high quality with limited options for drastic improvements of their final performance in the pots (high anode density, high current efficiency in pots and finally the low anode net consumption).

Nevertheless, a first objective identified is to reduce anode density variability at the anode paste plant, thanks to the predictive functions that will be developed during MONSOON iteration #1 and improved in iteration #2.

The first reason of anode density variability at the anode paste plant is the quality of raw materials. This point cannot be addressed within the MONSOON Project.

So the focus will be put on the second root cause of anode density variability that is equipment stoppage and start-up, in particular the mixing and forming chains. The important notice is that the failure of any part of the chain (upstream or downstream in the Paste plant) means a full stop of the whole chain.

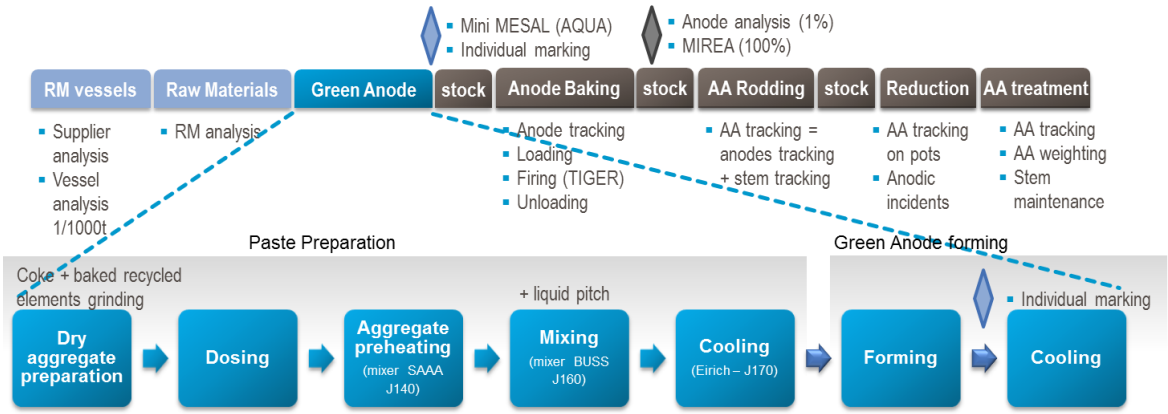

Figure 1 - Green anode production process

The mixer is key equipment for the paste production at the paste plant which is stopped in case of breakdown.
There is no redundancy on the mixer mainly because of the high initial CAPEX cost of such equipment.

The forming process is realized by two vibrocompactors. For safety reasons, it is impossible to operate only one machine, so maintenance operations are done during Paste plant shutdowns.

Nowadays, during the regular quarterly maintenance shutdowns of the anode production line, the process equipment's are checked and maintained as instructed by the equipment suppliers (defective pieces replacement and preventive maintenance actions). This implies a significant risk of non-detected trend towards a catastrophic failure, synonymous with longer maintenance stops.

2.1.2.1.2 Aim – Desired End State and Benefits

The objective is to anticipate the breakdowns that impact final green anode quality (e.g. anode density). The predictive maintenance will finally allow convenient scheduling of preventive maintenance, including optimization of resources (human and materials), and to prevent unexpected equipment failures and machine downtimes, and maybe higher system reliability. The additional possible outcomes will be enhanced safety management, optimized spare parts inventory management, increased equipment lifetime and improved energy consumption.

Furthermore, as a result of equipment failure, the rejection rate increases until a stabilized Paste Plant operation is reached. The Preventive Maintenance will help to detect critical situations early before a catastrophic failure, avoiding Paste Plant scraps.
For this purpose, existing measurements on equipment should be used only (voltage, current signals, etc.).

Initially, it isproposed to work incrementally, following this arrangement:

- POC #1.1: Mixer BUSS J160
- POC #1.2: Mixer BUSS J160 + Vibrocompactors K40/50
- POC #1.3: All J+K chains (+related H, M, G)

Note: during the first iteration, the working on the BUSS Mixer was started and a potential in studying the EIRICH Cooler (J170) could be seen.

2.1.2.1.3 Measurement

The definition of the predictive functions will be led by 3 KPIs: detection, anticipation and precision, in order to obtain robust alerts.

The final outcomes of this use case will take the shape of alerts or detailed actions to guide operators and supervisors to realize adapted actions.

For predictive maintenance, the final information displayed will be immediate / planned maintenance on the equipment (the development of the alert visualization is part of a parallel project). Additional contextual information will be provided to maintenance experts to organize and prioritize maintenance activities.

2.1.2.2 Equipment misbehaviour (POC2)

There are similarities between POC1 and POC2. In the next sub-section, only the particularities of the POC2 are depicted.

2.1.2.2.1 Problem Statement

The problem statement is similar to the equipment stoppage description. The specific point is that the anode density variability can also be impacted by equipment deviation or misbehaviour. During the first iteration, thefocus had to be put on the BUSS Mixer and the EIRICH Cooler. Several discussions with process experts from Aluminium Dunkerque led to analyse at the same time several key equipment of the production chain, including the BUSS Mixer and EIRICH Cooler, as well the condenser, the vibrocompactors, and several heating equipment,

Deviation to the nominal running settings of these equipment can have a major impact on anode quality.
The "mixing intensity" is the capacity of these machines to produce, as fully as possible, the following state of the green anode forming material:

- Dry product grains coated with binder,
- Dry product grain porosity filled with binder,
- Intergranular space minimized and filled with binder,
- Appropriate temperature for the forming step.

According to experts' knowledge, the following variables should particularly includein the input data (the corresponding thresholds are given in appendices):

- BUSS Mixer
    - Average amperage
    - Heat transfer fluid temperature
    - Pitch temperature
- EIRICH Cooler
    - Filling rate of the machine
    - Tool amperage
    - Paste temperature
- Vibrocompactors
    - Paste weight in the transferring hopper

Nowadays, like for the equipment stoppage scenario, the process equipment are checked and maintained during the regular quarterly maintenance shutdowns, with the same potential critical impacts.

2.1.2.2.2 Aim – Desired End State and Benefits

The objective of Equipment Misbehaviour analysis is to determine the different behaviours of the key equipment. Once the behaviours identified, the possible correlations between a given behaviour and a deviation and decrease of the anode quality can be found and give information of a trend to maintenance and process experts, in order that they can decide to adjust equipment or process parameters (where possible) or to plan maintenance (if needed considering the obtained variables).Unknown behaviours, that do not directly impact the anode density, but for which actions should still be taken in order to get back to standard process conditions could also be discovered. In online mode, alerts can also be arisen when unknown behaviours arise.

The approach for the POC 2 is called "unsupervised". The algorithms will group together the data in different classes without any *data-scientist* inputs to help him. The idea is to see if possible unknown behaviours can be found.

The ultimate goal of Equipment Misbehaviour analysis is to improve the overall anode quality.
As for the equipment stoppage scenario, enhanced safety management, optimized spare parts inventory management, increased equipment lifetime and improved energy consumption are expected.

Incremental work is proposed, following this arrangement:

- POC #2.1: Mixer BUSS J160
- POC #2.2: Mixer BUSS J160 + EIRICH Cooler + Condenser
- POC #2.3: All J+K chains (+related H, M, G)

2.1.2.2.3 Measurement

The outcome of this predictive function will be a list of behaviours. Those behaviours will have to be annotated by process experts. The predictive function is validated if the identified behaviours make sense in a business sense.
The final outcomes of this use case will take the shape of alerts or detailed actions to guide operators and supervisors to realize adapted actions, in order to go back to the best possible behaviour.

2.1.2.3 Prediction of the anode quality (POC3)

There are similarities between POC2 and POC3. In the next sub-section, only the particularities of the POC3are depicted.

2.1.2.3.1 Problem Statement

The problem statement is similar to the equipment misbehaviour description. The equipment misbehaviour detected within POC2 can lead to anode defaults that impact pot stability and performance. Some relations between equipment settings variation, raw material recipe and anode quality are already known. But some are not yet discovered and need to be controlled. Combinations of effects are also suspected.

For the POC 3, the idea is to train a supervised algorithm to recognize the conditions that lead to lower anode density.

Nowadays, when a variation is discovered (sometimes several days after the whole anode production process is ended), adjustments are made on:

- Raw material recipe (content of pitch and fines)
- Equipment settings

Both strategies can be leveraged in order to correct the process and meet the quality criteria for the anodes.

2.1.2.3.2 Aim – Desired End State and Benefits

The first objective is to model the anode density(green anode for iteration #1), using several process parameters, and then validatingthis approach on real time data at the batch level.

In a second phase, green anode quality at the individual anode level (iteration #2) should be predicted.
The second objective is to recommend countermeasures (e.g. adjust parameters settings) to improve the anode quality.

By knowing the individual anode quality, anode defects on potscan be anticipated and reduce the exposition of operator to potentially hazardous conditions. In fact, less scrapscould also be expected by anticipating process deviations.
For this purpose, existing measurements on equipment only should be used (voltage, current signals, etc.).

Considering internal events, it isproposed to work incrementally:

- POC #3.1: Batch analysis – Modelling & Recommendation
- POC #3.2: Individual anode data analysis

Note: Contrary to what was firstly expected, and because of a suspension of internal project, ADisno longer able to retrieve data from the baking process and the Electrolysis process. The analysis is limited at the Paste Plant level.

2.1.2.3.3 Measurement

The key outcome will be the demonstration of the predictive function's efficiency (value confirmation) when plugged on real time operation on AD plant. The performance of the algorithms is also considered in terms of recalland precision, in order to obtain robust alerts.

The final outcomes of this use case will take the shape of alerts and detailed actions to guide operators and supervisors to realize adapted actions.

For anode quality, the final information displayed will be contextual information helping process experts to realize adjustments on the process parameters in the Paste Plant.

2.1.2.4 Electrolysis area scenario – Prediction of the bath height, of the pot thermal balance and of the anode effects (POC4)

2.1.2.4.1 Problem Statement

Nowadays, the bath liquid height and the pot temperature are measured only every 32h (in most smelters, based on the electrolysis organization cycle). Between two measures, there is no clear knowledge of these variables available.

These two variables have a great impact on the pot performances. These are the optimized energy consumption and current efficiency.

The illustration hereunder depicts the thermal balance of a pot. The major part of the energy consumed produces heat that is useful to maintain the temperature of the pot needed for the electrolysis reaction. Nevertheless, a large part of this heat is lost, and these losses should be kept in control.

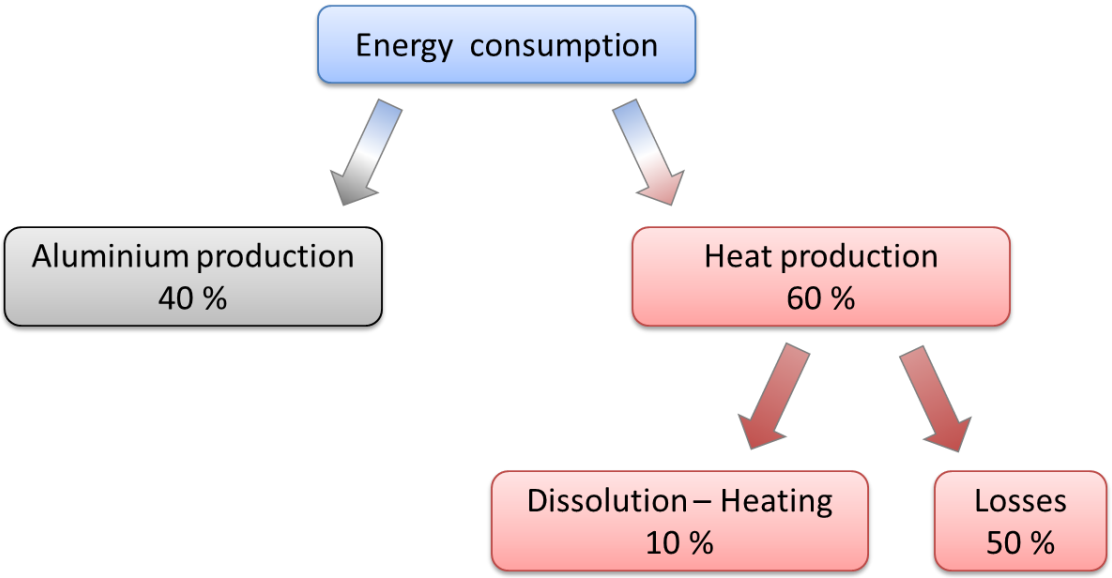

Figure 2- Breakdown of the energy consumption for the aluminium production, i.e. the thermal balance

The bath height is a key parameter. It is well known by electrolysis process experts that the bath volume impacts the alumina dissolution.

- If the bath height is too low the alumina dissolution will not be optimized that may lead to technical performance decrease (current efficiency).
- If the bath height is not well controlled, the bath may completely submerse the anode and may attack the pins of the anode, leading to metal pollution (with iron).
- If the bath height decreases, it may provoke the covering bath to collapse into the liquid bath and so modify the bath chemistry and the insulation properties.

The second aspect that can help managing the pot is its thermal balance. Indeed, the pot temperature must be kept between ranges in order to prevent:

- If the temperature rises, the bath chemistry changes that might lead to pot instability (and anode effects) and a decrease of aluminium production by re-oxidation of the aluminium
- If the temperature decreases, the cryolite contained in the bath may solidify preventing the alumina dissolution and creating sludge at the bottom of the pot (must be avoided to ensure a proper movement of the liquid, of the current flow and of the anodes when raised).

Theidentified objective is to predict the bath height and eventually the pot thermal status based on the information given by the chisel-bath contact system (one piece of equipment of the pot), thanks to the predictive functions that will be developed during MONSOON iteration #2.

And finally, one of the consequences of the pot thermal balance instability is to increase the quantity of solidified bath inside the pot and as a result close the opening made by the chisel in the crust, preventing the alumina to be delivered to the pot correctly. This underfeeding of alumina can create the anode effects that disturb the pot and its performances.

Nowadays, when a variation is discovered, sometimes after anode incidents or pot instability, adjustments are made on:

- Anode/Cathode distance (ACD), impacting the current efficiency
- Adjustments of the bath (volume and chemistry)

The overall scenario is foreseen as a process optimization scenario.

2.1.2.4.2 Aim – Desired End State and Benefits

The first objective is to give an estimation of the bath height contained in the pot, based on historic data, and then validates this approach on real time data. In a second phase, if possible, the pot thermal balance should be predictable.

The final target is to be able to give indications on the appropriate operations to be done on the pot (adjustments of the bath volume)

The second objective is to anticipate process deviations via predictive alerts and take countermeasures (e.g. adjust parameters settings) to improve pot stability.

Thanks to a better management of the bath height and pot thermal balance, the reach of higher current efficiency can be expected. An optimization of the *crustbreaker* usage (less consumables and less compressed air consumed) can also be expected.

By managing the pot thermal balance, theanticipation of anode effects and the reduce of the exposition of operator to potentially hazardous conditions should be possible.

It isproposed to work incrementally:

- POC #4.1: Bath height
- POC #4.2: Thermal balance
- POC #4.3: Anode effect prediction

2.1.2.4.3 Measurement

The key outcome will be the demonstration of the predictive function's efficiency (value confirmation) when plugged on real time operation on AD plant.

The final outcomes of this use case will take the shape of alerts or detailed actions to guide operators and supervisors to realize adapted actions.
The final information displayed will be immediate actions on the pot. Additional contextual information will be provided to process experts to realize adjustments on the process parameters on the pot or on the whole electrolysis hall.

2.2 Plastics domain

2.2.1 Description of the process to obtain the Engagement charters

For the plastic domain the concerns are focused on the quality of the production. The mindset of GLN is to improve the efficiency of the production and consequently good quality parts delivered to the Clients.

The definition of the POC is to orient the efforts of the MONSOON tool to improve the processes and methodologies to achieve a better quality of service.

Since the process depends on equipment's and tools, the POC is based on Injection Machines and Molds behaviour (with and without sensors) and the outcomes from the injection process (based on the quality control methodologies).

In recap of the selected scenarios, Business Case 1 was chosen to improve a very high cadence production, with a very mature process with some issues to be solved, regarding the mold maintenance and quality control results.For the Business Case 2, a low cadence production with very tight tolerances has been chosen, where the mold behaviour must be monitored in order to achieve good produced parts and less production stoppages.

In the Plastic domain, 5 different engagement charters are defined, distribute has 1 common POC for both Business Cases, and 2 for each one.

Machine Stoppages (POC1) – 3 sub scenarios:

- Krauss Maffei Machines
- Husky Machines
- Arburg Machine

- Business Case 1 – Coffee Capsules
  - Mold Maintenance without sensors (POC2)
  - Plastic Part Quality with automatic inspection (POC3)

- Second Business Case – Automotive Connector PSS1
  - Mold Maintenance and behaviour through sensors (POC4)
  - Plastic Part Quality based on manual inspection (POC5)

In the next sub-section, each scenario will be introduced taking as a basis the corresponding engagement charter.

2.2.2 Description of the scenarios and explanation regarding the planning

2.2.2.1 Machines stoppages (POC1)

2.2.2.1.1 Problem Statement

A machine stoppage can be described as an interruption of the injection molding process caused by a defect of the injection molding machine itself or by a defect of the injection molding process which leads to a stoppage due to the defined safety parameters.

Machine stoppages can be divided into expected machine stoppages and unexpected machine stoppages. Expected machine stoppages can be maintenance work, the end of a shift, etc…

Unexpected machine stoppages are mainly related to issues with the process, for example if a part sticks in the parting plane, if there is trouble with material or with the periphery of the injection molding machine.

2.2.2.1.2 Aim – Desired End State and Benefits

The main objective here is to minimize the time an injection molding machine stands still, due to unexpected errors.

After every machine stoppage approximately, the parts produced by the first 15 to 20 cycles are waste parts because the *plastification* of the plastic melt and the cooling system of the molding tool need some time to

reach an equilibrium regarding the temperature of the mold and the quality of the plastic melt. If the production runs on a 32 cavity mold approximately 600 parts are waste parts.

Furthermore, it takes several minutes up to one or more hours to fix the problem that caused the machine stoppage, so that the machine cannot produce parts during that time and cannot earn money.

2.2.2.1.3 Measurement

The KPI related to the amount of machine stoppages is to reduce the amount of unexpected machine stoppages to five stoppages per machine and day. The reaching of the KPI can be monitored with the help of the error tables being delivered from the injection molding machines from GLN.

2.2.2.2 Mold Maintenance without sensors (POC2)

2.2.2.2.1 Problem Statement

The POC 2 is related to the systematic production of waste parts. This can happen if there are breakdowns in the mold, which do not lead to a machine stoppage. This can happen if a pin breaks, which forms holes in the bottom of the capsules, or if the wearing of the mold is becoming too high, that for example the diameter of the ovality is not formed properly. The specific parts will then be rejected by the optical quality control system.

The POC 2 leads to a temporary rejection rate of 100 % from one specific cavity, because if the mold breaks or is damaged it will not repair itself, so the production of waste parts in that cavity will continue. The repairing of this must be done during a mold maintenance.

2.2.2.2.2 Aim – Desired End State and Benefits

From a technical point of view an alarm should be created if 30 parts from one cavity in a row are detected as waste parts due to the same error (mainly diameter, ovality, number of the holes).

Currently visual maintenance is done every shift (8h) to see if a damage of the mold is visible. If this is the case the machine worker can stop the process and close the affected cavity and restart the process with adjusted process parameters (i.e. dosing hub) in order to prevent the machine from producing a waste part each and every cycle.

Currently it is unclear since when the error occurred in the mold. Worst case would be that the error occurred directly after the last visual maintenance so that waste parts were produced during the whole last shift, which then lead to a significant increase in the amount of waste parts.

2.2.2.2.3 Measurement

In general, it can be assumed that the optical quality control system measured every defined quality criteria on the capsules and the lids and recognizes the cavity number of every capsule and every lid. The optical quality control system exports then the evaluation of the quality criteria. In addition, it can be assumed that the quality control for the PoC 2 is related to the quality criteria diameter, ovality and number of holes being delivered by the optical quality control system for the capsules. The related quality criteria for the lids are mainly the diameter of the lids and the height of the injection point.

2.2.2.3 Plastic Part Quality with automatic inspection (POC3)

2.2.2.3.1 Problem Statement

The POC 3 is related to coincidently production of waste parts which are not caused by a machine error or a mold break down. These waste parts can either be related to deviations in the process parameters, such as molding temperature, coolant temperature, injection speed or an inhomogeneous melt quality or also to coincident events, such as a capsule or a lid touches the side wall of a conveyor belt and receives a scratch from that.

2.2.2.3.2 Aim – Desired End State and Benefits

Here all defined quality criteria from the optical quality control system should be evaluated to decrease the amount of waste parts.

2.2.2.3.3 Measurement

The KPI to reach the POC 3 is to reduce the amount of waste parts under 2.25 % related to the amount of produced parts.

2.2.2.4 Mold Maintenance and behaviour through sensors (POC4)

2.2.2.4.1 Problem Statement

The POC 4 describes the mold behaviour. Similar to POC 2 systematic errors should be handled here by monitoring the values of the in-mould sensors and the machine parameters. A systematic misbehaviour from the injection molding machine or the injection molding tool leads to a rejection rate of 100 % which is detected by manual inspection in this business case.

A possible scenario can be the heating up of the mold over ling time scales due to a closure of the cooling channels, which leads to different temperatures in the mold and therefore to a deviation in the filling process. This can be detected with the help of the temperature sensors. Due to wearing of the mold it can also happen that the wearing in one side of the hot runner or the cold runner is higher than in the other side so that the connector parts are not filled equally. This can be detected with the help of the mold sensors.

A short shot can also be detected if this happens at the position of the temperature sensor. The temperature sensor has been positioned at the end of the flow path to ensure the complete filling of the mold.

2.2.2.4.2 Aim – Desired End State and Benefits

From a technical point of view an alarm should be created if 30 parts from one cavity in a row are detected as waste parts due to the same error (mainly short shots, deviations in the pressure or the temperature signals from the sensors).

2.2.2.4.3 Measurement

Currently the quality control is done manually by machine workers. The holes are proven by a gauge (jig) to ensure the correct diameter and the roundness of the holes. The complete filling is proven by a visual inspection. The machine worker looks if the connector parts were filled completely.

2.2.2.5 Plastic Part Quality based on manual inspection (POC5)

2.2.2.5.1 Problem Statement

The POC 5 is related to coincidently production of waste parts which are not caused by a machine error or a mold break down. These waste parts can either be related to deviations in the process parameters, such as

molding temperature, coolant temperature, injection speed or an inhomogeneous melt quality or also to coincident events.

2.2.2.5.2 Aim – Desired End State and Benefits

Here all defined quality criteria from the requirements list should be evaluated to decrease the amount of waste parts.

2.2.2.5.3 Measurement

The KPI to reach the POC 3 is to reduce the amount of waste parts under 3% related to the amount of produced parts.

3 Impact assessment (through PoV)

In this chapter the assessment of the Monsoon platform will be addressed both for aluminium and plastic domain. The specific KPIs will also be addressed, but as an introduction the base layer KPIs are presented because they are cross sectorial and can address any kind of industry.

The base layer is mainly composed of Environmental KPIs: this is a reasonable configuration, for the Environmental class shows high cross-sectoriality. The selected indicators have been chosen according to a reliability criterion, which favoured KPIs with high robustness concerning the impact assessment method and the characterization factors to be adopted. All the Environmental KPIs are computed by the Life Cycle Management plugin implemented within WP5 activities; further information concerning this component is available in **Error! Reference source not found.**.

The only process KPI included in the base layer is the process yield. This is because each process is expected to experience an increase of its yield due to the adoption of optimization techniques. Process yield is computed as a by-product of the LC plugin, as the evaluation of the yield is a required step in the process to compute Environmental KPIs. Those KPIs are presented in Table 1

Table 1 - Base layer KPIs

KPI name	Unit	Origin	Class	Description
Global Warming	kg CO_2 equivalent	LC management plugin	Environmental	Total amount of equivalent greenhouse gases generated by the investigated process
Primary Energy Consumption	MJ equivalent	LC management plugin	Environmental	Total amount of primary energy required to manufacture the investigated product
Direct Energy Consumption	MJ equivalent	LC management plugin	Environmental	Total amount of energy (electric and thermal) directly consumed by the investigated process
Electricity consumption	MJ	LC management plugin	Environmental	Total amount of electricity directly consumed by the investigated process
Raw material consumption	kg	LC management plugin	Environmental	Total amount of material required to manufacture a unit of valuable product
Recycled content	%	LC management plugin	Environmental - process	Percentage of recycled material in the investigated product
Water	l	LC management	Environmental	Total amount of water required to

Consumption		plugin		manufacture the investigated product
Waste to landfill	kg	LC management plugin	Environmental	Total amount of waste originated from the process which manufactures the investigated product and sent to landfill
Waste to recycling	kg	LC management plugin	Environmental	Total amount of waste originated from the process which manufactures the investigated product and sent to recycling
Process yield	%	LC management plugin	Process	Ratio between the valuable output of the investigated process and the total input material
Product Circularity Index	%	to be defined	Circularity	Grade of circularity of the investigated product along its whole life cycle

3.1 Aluminium domain

In a first paragraph the KPI for the aluminium domain will be reminded and the main ones for the carbon area and the electrolysis area will be highlighted. Then thanks to the evolution of those KPIs the associated financial gain at this current state of the project will be estimated.

3.1.1 Carbon area specific KPIs

For the first iteration, in the carbon area, with the focus on the green anode production, two different process KPIS have been identified:

Table 2 – Aluminium specific KPIs

KPI name	Unit	Origin	Class	Description
Anode quality	number of low quality batches	Dedicated predictive function	Process	Anode quality is evaluated referring to anode density; according to the computed density value, a batch of anodes is annotated as *good* or *bad*
Anode rejection rate	%	Dedicated predictive function	Process	Rejected anodes need to be recycled, with extra consumption of energy and machinery due to the crushing and re-forming processes

The variability of the green anode density affects their behaviour on the pots and can lead to instability of the pot and to anodic incidents.

Any rejected anode is reintroduced in the green anode production process. The anode blocks are crushed in a specific workshop to a required dimension before being reintroduced as green recycled product. Even if the impact of raw material is reduced thanks to the internal recycling process, this specific step is energy consuming and leads to inefficiencies along the production process.

For iteration 1, batches of anodes produced per 30 minutes periods are worked on. A period is considered of lower quality if at least one anode produced has a density below 1620. The KPI is therefore the number of such lower quality periods.

3.1.2 Electrolysis area specific KPIs

In the electrolysis area, study the impact of process parameters on the bath height and the pot thermal balance will be studied.

For this study, the following KPIs will then assessed:

- Bath height
- Temperature of the pot.

3.1.3 Impact assessment for the Carbon area

The impact of the Monsoon solution has been calculated for POC3. For POC 1 and POC 2 the results of the predictive functions are not enough relevant yet.

The estimation is separated by each type of anode defect. [RD.4]explains in detail those incidents, but as a reminder they can be described as followed:

- **Dusting in pots:** lower density of the anodes induces more carbon dusting, leading to more anodes defects (mushrooms) and more instability and human operations on the pot. The main effect of Monsoon is to reduce the low-density anodes, in order to maximise the production, reduce the human operations and reduce the instability of the pot.
- **Net carbon gain:** higher density induces less anode consumption and less carbon consumption. Again, the anodes are changed less frequently and there are less operations on pots.
- **Incidents on pots:** related to the decreases of anodes incidents on pots.
- **$C0_2$ emission:** less anode consumption induces less $C0_2$ emission.

The gain brought by the Monsoon project are calculated thanks to some hypothesis on the evolution of the anode effects:

- **Dusting in pots:**420 k€ / year
- **Net carbon gain:** 55 k€ / year
- **Incidents on pots:** 60 k€ / year
- **CO_2 emissions:** 60k€ / year

Overall the estimated gain is around 600 k€/year for the Carbon area only. This calculation has to be confirmed on a longer time period and will be then recalculated at the end of the project, with the impact assessment in the electrolysis area, and both will be put in deliverable D7.4 – Final Demonstrator in the Aluminium and Plastics Domains.

3.1.4 Impact assessment for the Electrolysis area

In this case the impact assessment is calculated for the POC 4 based on the gain detailed in the 2.1.2.4 paragraph. As a reminder we are hoping to increase the current efficiency of the pots for POC 4.1 and POC 4.2, and reduce the anode effect, thus the instability of the pot thanks to POC 4.3.

Associated to those improvements the estimated financial gain is around 2.5 M$ per year, which is considerable and very interesting for AP.

An environmental assessment based on LCA methodology will be carried out to demonstrate the environmental benefits of optimization of electrolysis control. This activity will share the same approach adopted for anode quality use case and described in this document in section **Error! Reference source not found.**.

At the present stage of the MONSOON project, general understanding of electrolysis use case has been performed. Involved partners, mainly LCEN, AP and CAP, are investigating specific data structure and availability to proper design the LCA plugin that will perform the environmental assessment. As the activity is currently running, no specific results are available for official reporting.

Results of LCA activity in the electrolysis area will be reported in D7.2 – Final Evaluation Framework and D 5.8 – Final Life Cycle Management plugin. Environmental KPIs will include for sure Global Warming and Primary Energy Demand, as the process is highly energy-intensive; additional KPIs are being investigated and discussed at present stage of the project.

3.2 Plastics domain

The KPIs are divided in two different blocks: a common layer where cross-sectorial KPIs are reported, and domain-specific KPIs, where indicators that are relevant just for some use cases are stored. All the details about those KPIs can be found in [RD.3]

For the plastic scenario, the PoV (Proof-of-Value) it is supported in two big pillars: a) the production management and b) the client's satisfaction, concerning the delivered end-products. Due to this alignment the PoV will be built on a) efficiency of the process/management and b) earnings from selling parts.

For the presented PoC's, the expected gains are correlated to direct and indirect earnings. It is expected that the reduction of stoppages and the waste of raw material (through NOK parts) will represent less effort on process, higher machine occupation and resource consumption reduction, to the same planned production quantities. In another vector, the reduction of NOK parts, will increase the direct earnings on the product selling and decrease of client claims.

3.2.1 Coffee Capsules Specific KPIs

For the first iteration, in the Coffee Capsules production, with the focus on the quality of the plastic parts, one process KPI has been identified:

Table 3 – Plastic specific KPIs

KPI name	Unit	Origin	Class	Description
rejection rate	%	Dedicated predictive function	Process	Percentage of waste capsules produced

The plastic injection process is a very important factor for the quality of the produced parts, along with other factors, such as the quality of the raw material, the production area temperature, behaviour and wear of the mold.

In the actual scenario the rejected parts are not being introduced in the line (through a recycle process), because of client limitations. Based on this the overheads related to the raw material consumption will reflect how solid is the production management.

Although of having a rejection rate between 1,5% to 2%, this can be improved by applying new predictive methodologies and a closer analysis of the causes. The KPI will be calculated from the rejection rates, based on predictive maintenance along with injection process monitorization and trend analysis.

3.2.2 Impact assessment for the Coffee Capsules Production

For the Coffee Capsules production, it is expected to increase 5-8% of the earnings from the minimization of the machine stoppages and reduction of raw material consumption wasted on NOK parts production.

The increase of the efficiency of the process will represent a reduction of machine usage (assuring the client demands); indirectly this will represent a reduction of energy consumption approximately around 20%, based on the required working days to complete the orders: 7 to 5 days/week.

Because the production line doesn't have yet a year reference it is not possible to quantify the earnings in terms of €/year; however, it is expected to reach savings around 100k€/year based on the energy consumption management and waste production/raw material consumption

3.2.3 Impact assessment for the Automotive Plastic Parts

In the case of the Automotive Parts, the biggest concern is the quality of the plastic part, due to the type of part – safety connector for Automotive Car. With the monitorization of the mold through sensors, the injection process will be improved and the defects in the end-product will be reduced. This will represent an increase of earnings around 8-10%.

The mold maintenance it's also a key topic: with a good control of the behaviour of the mold, the maintenance plan will be more accurate and functional. With this, it will possible to reduce the maintenance costs, and increase the sales earnings up to 12%.

However, this estimation will be updated in the final version document – D7.4 – Final Demonstrator in the Aluminium and Plastics Domain, as described also on the Aluminium Domain chapter.

4 Exploitation of the results of the demonstrator

4.1 Aluminium domain

4.1.1 Ongoing activities for Equipment stoppage, Equipment deviation and Pot behaviour

4.1.1.1 Progress on Equipment stoppage (POC1)

4.1.1.1.1 Problem definition details

As mentioned above, particular attention has been paid to the BUSS mixer (J160) and the cooler (J170) stoppages, because these two types of equipment are the most critical for the anode production. It is desired that a stoppage is forecast about 45 minutes before its occurrence. For this purpose, process and equipment data from these equipment types are inserted (after pre-processing) as input to a supervised or unsupervised machine learning binary classification model, which determines the chance of an upcoming fault by providing an anomaly score together with an alert when needed (abnormal class).

For training, first a file with stop information is used. This file records stops and their periods in two different ways:

1. According to the timestamps registered on the internal system, which also records the fault types.
2. According to the following misbehaviour of any of the next variables (which does not specify the fault type), as has been proposed by Rio Tinto Aluminium experts:
    - J160_INT_MOY_MALAXEUR[1]<200
    - J010_VIT_MOT_VIS_DEMANDEE[2]< 50
    - J050_DEB_INSTANTANE_DOSEUR[3]< 1

These three variables will be called "stop annotators".

In theanalysis, it has been assumed that a time point corresponds to a fault existence if a fault is recorded by any of the above two ways. When a fault period defined by the second way begins at a moment when no fault is mentioned yet by the firstone but is mentioned later and before the fault period defined by the second way ends, then it is assumed that a fault linked to the equipment specified by the system (first way) starts at the first abnormal time point of the second way. Usually, such a fault is identified by the second way up to about 10 minutes prior to the first. Thus, in the following, when writing expressions like "a fault is predicted *n* before its occurrence", this few-minute difference is not included in *n*. In other cases, there are fault periods recorded only by the first way. The type of these faults isconsidered as undefined.Figure 3describes such kind of classification. Stop information is provided in both ways in convenient files from September 2016 to February 2018, thus this period is the one analysed at this stage. Only data from operation periods have been taken into account for the classification model.

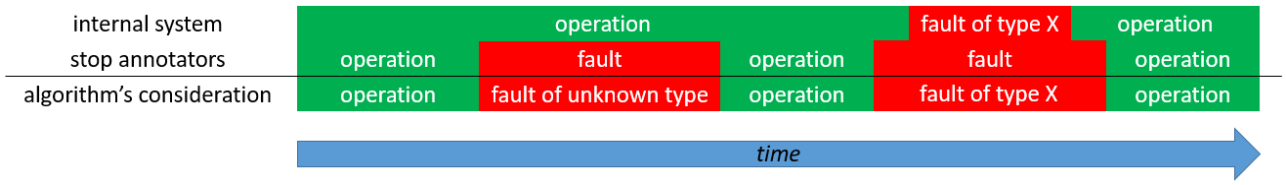

Figure 3 - Algorithm's consideration for fault periods and their types, based on information of the internal system and the stop annotators.

[1]J160 Buss mixer average intensity
[2]J010 dosimeter screw requested velocity (under very coarse silo)
[3]J050 dosimeter instant flow (fine silo)

The variables that have been characterized as directly related to the BUSS mixer and the cooler stops are those from the J16X(J160-J167) and J17X (J170-J178)series respectively. The data analysis aims at extracting the variables (or features of them) that can timely forecast a stoppage of the respective type.From these chains, process (continuous) data are available within the whole 18-month period of analysis, whereas equipment data (usually binary variables indicating normality/abnormality) are available since July 2017.

The initial time step of the data was from 2 sec to 5sec.Since a constant time step is needed for classification, the data have been aggregated with a 10-sec step. When running the code locally, no memory constraints hinder the code from running. In the near future, all these data will be extracted with the help of the *DataLab* instead of manually extracted files. Then, the matter of aggregation frequency will be re-examined. Unluckily some process signals have a frequency of about 1Hz, which cannot be identified.

Providing further mathematical details of the methodology is out of scope of this deliverable and will be documented in "D5.4: Final Online and Deep Machine Learning Functions" by July 2019.

4.1.1.1.2 Basic results and observations:

Within the 18-month period of analysis, there is a total operation interval of about 8.14 months. Totally, 980 fault incidents immediately after normal periods were registered by any of the two aforementioned ways. These are divided into 427 incidents which are identified only from the stop annotators and 553 that are recorded by the system, 25 and 26 of which are registered as related to the cooler and the BUSS mixer respectively.

There are data from 10 process variables from each of the J16X and J17X series within the entire interval, as well as more than 100 equipment variables from each series since July 2017.

Most variables (even variables not included in the J16X or J17X series) are influenced by the faults in the critical equipment kinds of the BUSS mixer and the cooler. However, it seems generally difficult to identify variables being able to forecast the faults. Usually, the behaviour of variables does not change before the faults have been detected by the stop annotators, even if it changes before the faults are registered by the system. Particularly, it needs to be mentioned that most equipment variables are constant or almost constant during the whole set of operation time points, although they often alter during faults. It seems that their behaviour is the consequence and not the cause of breakdowns.

Cooler: The most appropriate variable for timely forecast of cooler faults is J170_NIVEAU_MELANGEUR_REFROIDIS (J170 cooler paste level), because the increase in its variance is obvious.

Actually, during the last moments of operation before many cooler faults associated with high level in the cooler (according to the experts' and operators' comments), this variable gradually starts vibrating with a period of about 2min, so similar results are also obtained with a 1-min time step as well. The behaviour of other variables also changes before cooler faults, but not so timely, and usually not before the behaviour of stop annotatorsalters. The cooler faults with comment "high level in the cooler" have increased since July 2017 and especially in the beginning of 2018. What is more, the above variable seems to have caused many other incidents as well, most of which are identified only by the stop annotators. According to the above, for the cooler it has been decided to pay particular attention to the last part of the period of analysis, from July2017 to March 2018, within which 11 cooler faults with operation before have been registered (i.e. almost half the total number of cooler faults within the entire period of analysis), all of which are related to high level in the cooler, and other 62 faults have been identified by the stop annotators. (Of course, there are also other known types of registered breakdowns.). The interesting results discussed about the variable J170_NIVEAU_MELANGEUR_REFROIDIS are depicted in Figure 4,

Table 4 and Figure 5.

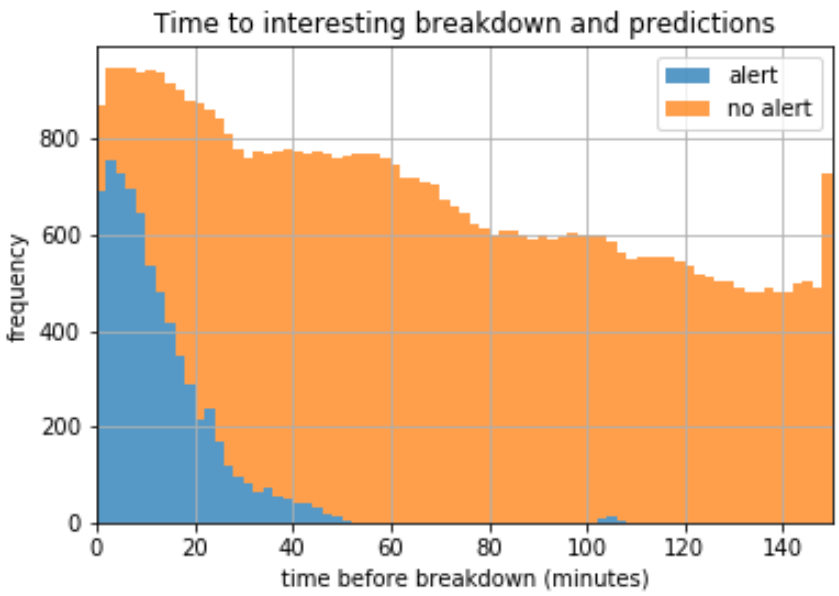

Figure 4 - Stacked[4] histogram of fault alerts' presence/absence as a function of time before breakdown, for J170, J150, J020, K030, L (stations d' arrêt – stop stations) and stops of undefined type that occurred from July2017 to March 2018, with the standard deviation of the variable J170_NIVEAU_MELANGEUR_REFROIDIS within a 1-min window as input. The correct percentage of outliers (~4,338%) is considered as known.

Most fault incidents have been forecast by the classifier's alerts within the last 15min of operation, whereas almost no alert is raised when time to the next breakdown exceeds 1h.

Table 4 - Classification confusion matrix and evaluation measures for J170, J150, J020, K030, L (stations d' arrêt – stop stations)[5] and stops of undefined type that occurred from June 2017 to March 2018, with the standard deviation of the variable J170_NIVEAU_MELANGEUR_REFROIDIS within a 2-min window as input.

CONFUSION MATRIX	Predicted normality	Predicted abnormality
Actual normality	153519	2285
Actual abnormality	2285	4781
EVALUATION MEASURES	**by trained model**	**neutral**
Precision	98.53%	95.66%
Recall	98.53%	100%
F-measure	98.53%	97.78%
Accuracy	97.19%	95.66%
Matthews Correlation	0.662	0

[4]Note that the tops of the orange bars correspond to the number of all observations, with or without alert.
[5]The reason to choose the J150, J020, K030 and L stops was an abnormality of J170_NIVEAU_MELANGEUR_REFROIDIS was found before their occurrence within the 2 first months of 2018. Since only 1 incident of each such fault type that is correlated with the abnormality of the above signal is recorded during this period, maybe these types are incorrectly registered, or the respective breakdowns almost coincide (by chance or not) with others, which are related to the cooler.

During operation periods the actual class has been defined as 1 (abnormal) when a fault begins at most 15min later, and as 0 (normal) otherwise. The correct percentage of outliers (~4,338%) is considered as known, thus the two types of wrong predictions have the same frequency. The neutral results are those that would be obtained by a model always predicting the most frequent (i.e. the normal) class. Common measures like precision, recall, F-measure and accuracy depend only on a part of the confusion matrix and/or are affected by the imbalance of the two classes.

For this reason, the most appropriate measure to evaluate the confusion matrix is the Matthews correlation, which acts like balanced accuracy (i.e. does not consider the classes' proportion) and ranges from -1 (completely bad) to 1 (perfect). This score is 0.662 for the above interval, whereas for the whole period of analysis (1/9/2016-28/2/2018) it would be only 0.1394, because most faults of interest of the previous months are not related to a high level in the cooler, even if they are linked to the cooler in general.

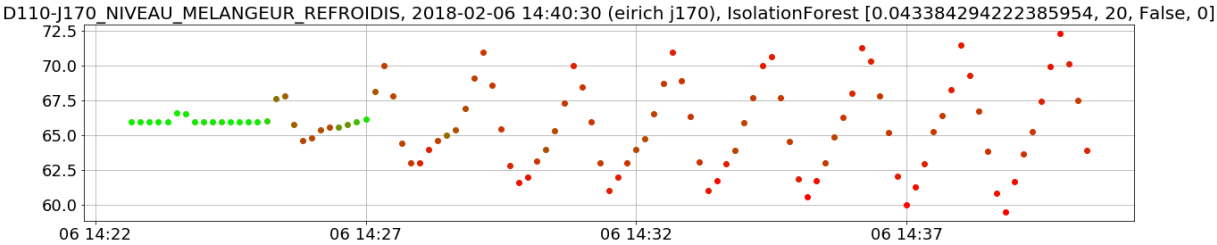

Figure 5- Representative [for most J170, J150, J020, K030, L (stations d' arrêt – stop stations) and type-undefined incidents in the first 2 months of 2018] example of the behaviour of the variable J170_NIVEAU_MELANGEUR_REFROIDIS during the last operation minutes before a cooler fault (right endpoint of plot's time interval) begins.

About 15 min before the fault this variable gradually starts vibrating (with a period of about 2 min), and as a result the classifier's anomaly score increases (hence the gradual change of the points' colour from green to red).

BUSS mixer: Unfortunately, for the BUSS mixer nothing is promising for short-term forecasting, either by naked eye or by a classifier. That is, no common change of behaviour of some signal(s) is detected among different incidents, either related or irrelevant to the mixer. There is no indication that a more sophisticated pre-processing or classification approach can actually help (although some trials have already been executed), because no timely change is observed in the data behaviour before the BUSS mixer faults. It is not guaranteed that other variables may be utilized to forecast such stoppages in the next step of the analysis. However, it is indisputable that very few BUSS mixer faults have occurred after March 2017 within the period of analysis. Maybe this can be explained by some configuration change. The only thing that can be stated is that many faults until March 2017 were linked by the experts to the equipment variable J164_T0_HUILE_TRES_HAUTE ("Oil temperature too high"), which did not exist until recently (June 2017).

4.1.1.1.3 Future work

As mentioned above, shortly all possible input data will be retrievable from the *DataLab*. This will enable real-time forecasting based on some pre-trained model and visual analytics in the MONSOON platform. New trainings will be executed when requested. Since the stop information is not provided in real time, training will require that this information is acquired partly manually for the training period.
The discussions with process experts are still on going, in order to interpret the signals' behaviour more thoroughly, and try to find other ways of forecasting interesting faults.

4.1.1.2 Progress on Equipment deviation (POC2)

4.1.1.2.1 Description of the approach

The approach developed for this POC is an unsupervised machine learning approach. The approach can be split in several steps:

- Selection (by process experts) of the most important process parameters for each equipment of interest
- Preparation (by data scientists) of those data – cleaning, pre-processing
- Application of unsupervised machine learning algorithms ("clustering techniques") on those data
- Presentation of the clusters / behaviours to process experts

At that point, either the clusters make a business sense to process experts – then move to the next steps (see below). If not, thereshould be a go back to the previous steps: either selecting other variables, applying other preparations, or other clustering techniques.

Once the obtained clusters are satisfactory for the process experts, it is necessary to:

- Train another model for recognizing those clusters in real-time
- Be able to raise alerts if new unknown behaviours arise in real-time

4.1.1.2.2 Results obtained so far

Several iterations of the approach were made in the first iteration of the project. the firstapproach on the BUSS Mixer is applied, and then on the EIRICH cooler and the Condenser. After several discussions with process experts from Aluminium Dunkerque, the decision was made to apply the approach to the entire J+K chain, e.g. the context of POC 2.3. Here only the results related to this last iteration are presented.

The list of most important parameters defined by process experts was the following:

D110-G130_DEB_BRAI_CORIOLIS_G130
D110-G130_TEMP_BRAI_CORIOLIS_G130
D110-G131_DEB_BRAI_CORIOLIS_G131
D110-G131_TEMP_BRAI_CORIOLIS_G131
D110-H070_TEMP_F_T_VIS_PRECHAUF_COKE
D110-H080_TEMP_F_T_MALAXEUR
D110-J140_TEMP_COKE_ENTREE_J150
D110-J140_VIT_ROTATION_VIS_PRECH
D110-J160_INT_MOY_MALAXEUR
D110-J160_MES_OUV_CLAPETS_MALAXEUR
D110-J160_TEMP_PATE
D110-J170_NIVEAU_MELANGEUR_REFROIDIS
D110-J170_TEMP_INSTANT_PATE
D110-J172_INT_MOT_TOURBILLON
D110-J173_DEB_EAU_REFROIDISSEMENT_J173
D110-J173_TEMP_EAU_REFROIDIS
D110-J173_VIT_POMPE_A_EAU
D110-K030_POIDS_BRUT_TREMIE
D110-K030_POIDS_NET_TREMIE
D110-K040_POIDS_PATE
D110-K050_POIDS_PATE
D110-M150_DEPRESSION_CONDENSEUR
D110-M150_To_EAU_ENTREE_CONDENSEUR

D110-M150_To_EAU_SORTIE_CONDENSEUR
D110-M150_To_VAPEUR_ENTREE_CONDENSEUR
D110-M150_To_VAPEUR_SORTIE_CONDENSEUR

Those 27 process parameters are related to the dosing of the raw materials, to the BUSS Mixer, the EIRICH cooler and their raw material preheating, as well as to the vibrocompactors and the condenser.

All those parameters are measured every second. In this approach, the elementary mesh of analysis is a period of five minutes of production. The preparation of the data is therefore the following:

- Aggregation of the data to 5 minutes periods
- Computation of simple statistical indicators to summarize each parameter's behaviour in the period: median value, standard deviation, 5% and 95% percentiles
- Suppression of periods for which the paste plant was stopped.

For this analysis, the data from September 2016 to June 2017 were firstly used. The prepared dataset contains approximatively 40000 periods of 5 minutes.

Once those data prepared, an unsupervised machine learning approach was applied in order to group them into different clusters / behaviours. The approach chosen is called "agglomerative hierarchical clustering". The idea is to start by associating each period to its own cluster. Then, clusters that are the most similar (in terms of a distance metric, here Euclidean) are merged together. Ahierarchy of clusters that can be represented by a tree (or dendrogram) is obtained.

The dendrogram obtained in the analysis is shown in **Error! Reference source not found.**.

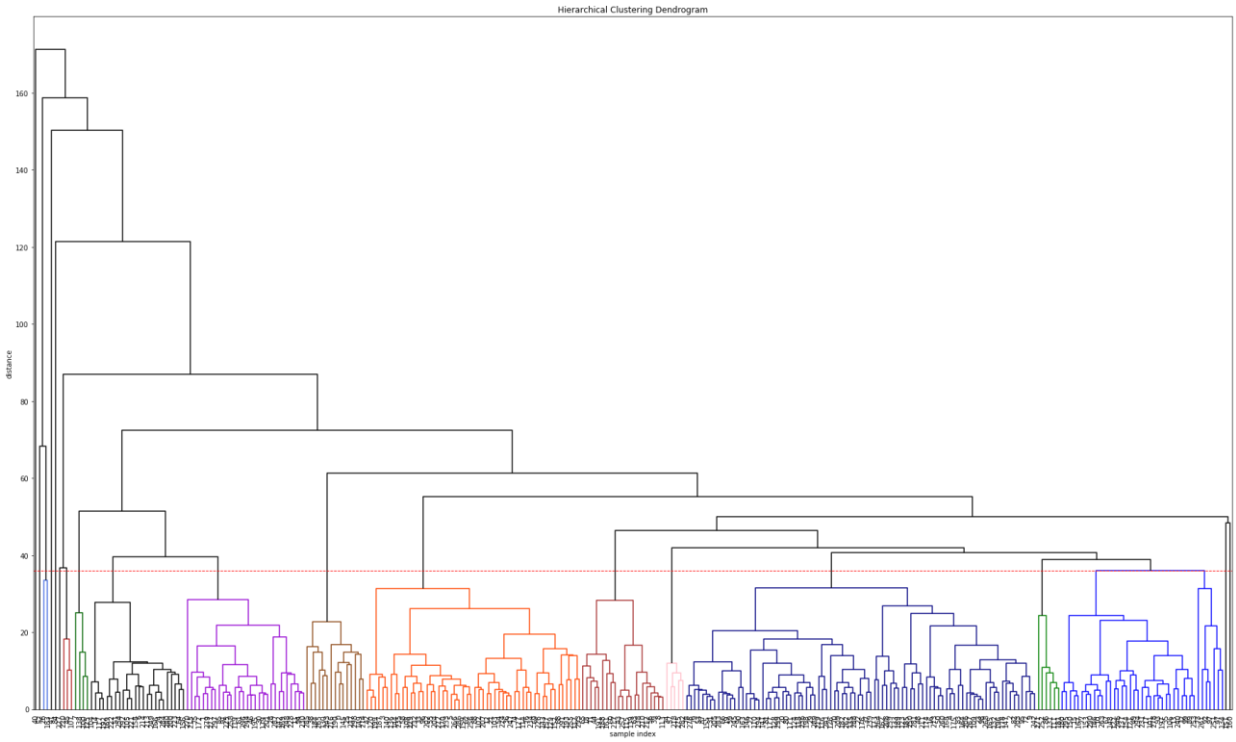

Figure 6 - Dendrogram obtained for POC 2

A threshold on the distance (y axis of the dendrogram) is selected, and all the periods belonging to the same branch below this threshold are associated to a single cluster.

Here, the result are19 clusters in the end. However, 8 of those clusters were containing less than 10 periods and were removed for the rest of the analysis. The parameters associated to those rare clusters are likely to be outliers that do not characterize in a realistic way the different behaviours of the paste plant.

Once the clusters built, they need to be analysed in order to understand if they make business sense. To do that,*data-scientists*looks at:

- The distributions of the different process parameters for each cluster
- The time repartition of those clusters

For the first point, the boxplots of each process parameter for each cluster are drawn, in order to compare all of them. For illustration purpose, those boxplots for 3 process parameters are shown: D110-J173_TEMP_EAU_REFROIDIS (Water cooling temperature of the EIRICH Cooler, **Error! Reference source not found.**); D110-H070_TEMP_F_T_VIS_PRECHAUF_COKE (Temperature of the BUSS Mixer pre-heating screw, **Error! Reference source not found.**); D110-J170_NIVEAU_MELANGEUR_REFROIDIS (Level in the EIRICH Cooler, **Error! Reference source not found.**). It canbe seen for instance from these plots that particular behaviours can be associated to some of those clusters. For instance, clusters 6 and 7 show both a higher D110-J173_TEMP_EAU_REFROIDIS and D110-H070_TEMP_F_T_VIS_PRECHAUF_COKE; while cluster 14 shows a lower-than-usual D110-J170_NIVEAU_MELANGEUR_REFROIDIS.

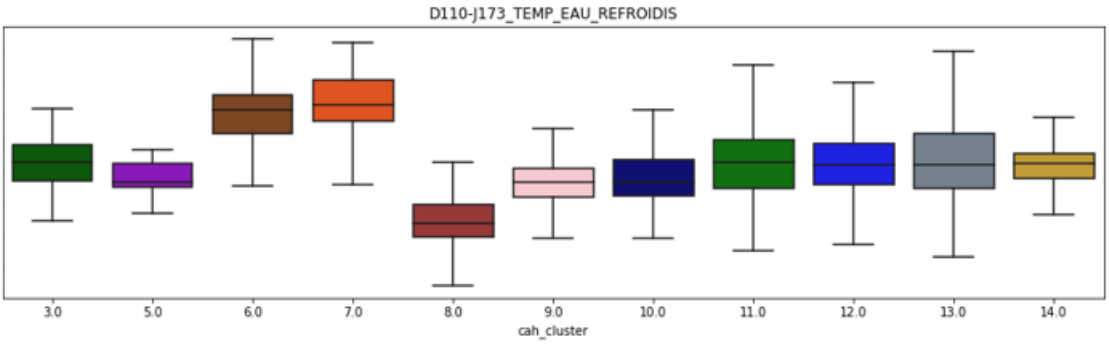

Figure 7 - Boxplot of D110-J173_TEMP_EAU_REFROIDIS for the different clusters obtained for POC 2. The values on the Y axis are hidden for confidentiality purpose

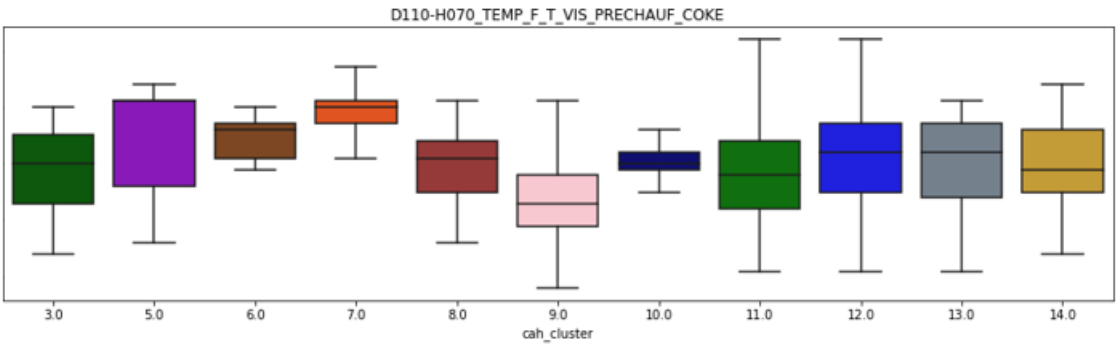

Figure 8 - Boxplot of D110-H070_TEMP_F_T_VIS_PRECHAUF_COKE for the different clusters obtained for POC 2. The values on the Y axis are hidden for confidentiality purpose

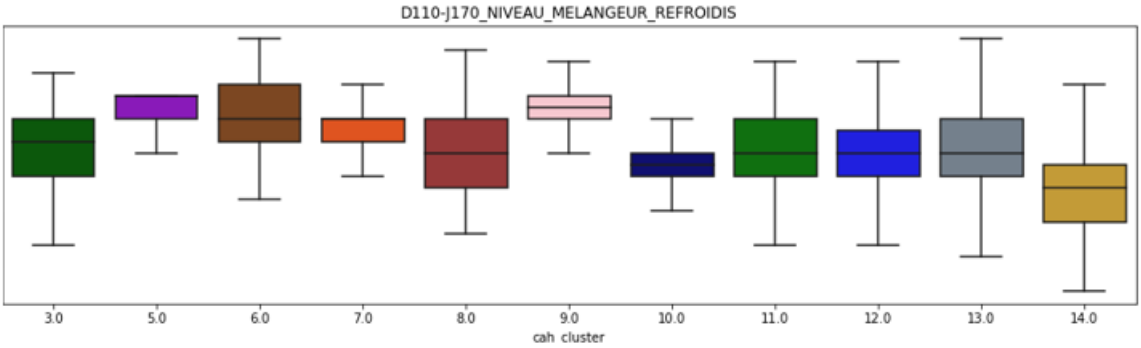

Figure 9 - Boxplot of D110-J170_NIVEAU_MELANGEUR_REFROIDIS for the different clusters obtained for POC 2. The values on the Y axis are hidden for confidentiality purpose

For understanding the time repartition of those clusters, their repartition for each month of data which were analysed was looked at.

This repartition is shown for the first three months of data analysed here, i.e. September to November 2016, in **Error! Reference source not found.**. Cluster 6 appears for instance only in September 2016, while cluster 14 is highly dominant in October 2016, before appearing much less frequently in the following months.

These results were discussed with Aluminium Dunkerque. However, beyond the interpretations of those clusters, and the definition of possible actions to mitigate abnormal behaviours, difficulties arose when trying to detect these clusters in a more recent dataset.

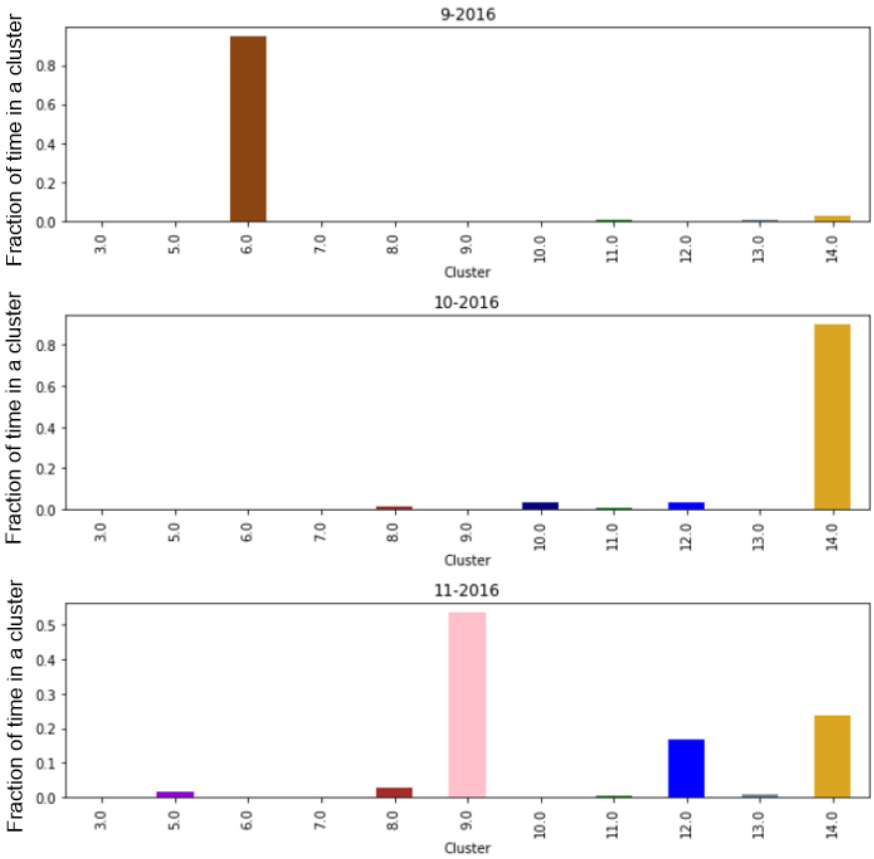

Figure 10 - Repartition of the different clusters obtained for POC 2 for the first three months of analysed data

4.1.1.2.3 Difficulties of the approach

As discussed in the previous sections, the goal of this approach is to be able to recognize in real-time if a new 5-minutes period belongs to one of the identified clusters; give a possible action if the cluster is abnormal; and raise an alert if an unknown behaviour appears.

To do so, once the clusters defined (see previous section), two machine learning classifiers are trained:

- A One-class Support-Vector Machine (SVM), which aims at detecting newcomers, i.e. determine if a period is similar to the periods used to create the clusters, or if it is a new behaviour;
- A Gradient Boosted Decision Trees classifier, which aims at associating a period to a given cluster, in case it is declared as a "standard" period by the one-class SVM.

Those models are trained on the same dataset which was used for the clustering, i.e. from September 2016 to June 2017. To test them, a new set of data is employed, from February 2018 to August 2018.

When applying the one-class SVM to this new dataset, all the new periods were tagged as **new behaviours**. This means that some of the process parameters that were used to define the clusters have drastically changed from June 2017 to February 2018. Thiscan be clearly seen for instance on the process parameter D110-H070_TEMP_F_T_VIS_PRECHAUF_COKE, whose distribution is shown in **Error! Reference source not found.**. The orange histogram shows the distribution of this parameter in the data used to train the clustering, while the blue distribution represents the data on the February 2018 to August 2018 period.

As some of the parameters used to define the clusters have drastically changed, the one-class SVM detects all the new data as new behaviours. This means that the clustering needs to be retrained on the latest dataset before being able to use the model in production.

The difficulty stands in the fact that automatizing a clustering approach is a highly challenging task. The analysis done here showed that some of the paste plant process parameters can highly change over time – an issue that was not obvious to Aluminium Dunkerque process experts.

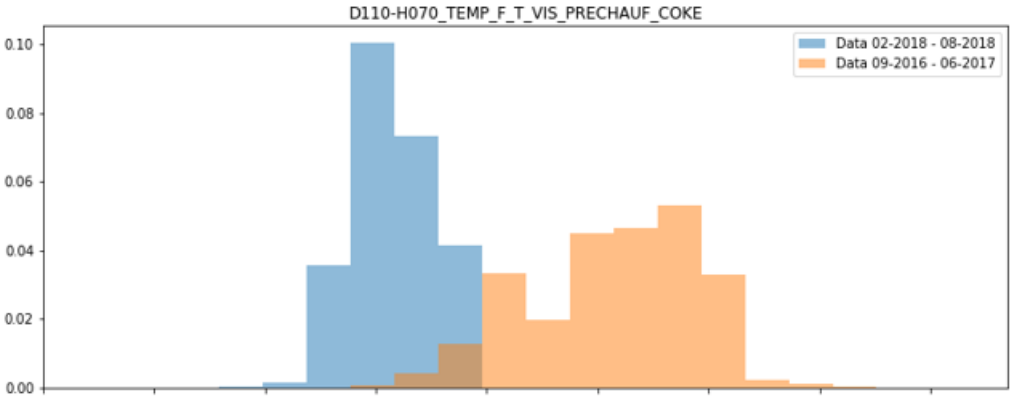

Figure 11 - Distributions of D110-H070_TEMP_F_T_VIS_PRECHAUF_COKE for the period September 2016 – June 2017 (orange) and for the period February 2018 – August 2018 (blue). The values taken by the parameter (X-axis) are hidden for confidentiality purpose.

For this reason, it has been decided to put on hold the POC 2, focusing the effort on the POC 3, which shows a more direct impact on anode density as well as an easier possibility of automatic re-training.

4.1.1.3 Progress on Pot behaviour (POC4)

4.1.1.3.1 Introduction

In the current section, an abstract overview of the methodology followed to control the Pot behaviour in the Aluminium domain is presented. Firstly, it will be discussed what variables were used to train our models and the way the given variables are modelled within the temporal framework that is posed by the sampling action itself. Subsequently, the algorithms and techniques that were used to develop the predictive models will be mentioned. Finally, some results and figures of the three best predictive models will be presented.

4.1.1.3.2 Feature selection

The same analysis is followed for the Bath height variable (HBMES or HB), the Metal height variable (HM) and the Thermal balance variable (TB). The aforementioned variables are sampled for each Pot every 32 hours. Around 220 extra variables that are sampled every 8 hours and themodels are based on them to find relationships between them are available, as well as the time and the variable that should be forecasted (e.g TB). After dropping the blank variables (those that did not have registered values) and passing the remaining variables(~180) through a pre-processing pipeline of feature selection techniques, it was determined that all the variables must be used, since all the models perform better using the whole set of variables. From now on, the referring to the HB, HM and TB variables as targets and to the rest variables as features will be done.

4.1.1.3.3 Modelling approach

The first approach and the naivest, but cheapest by means of computational time, was to use a window of past values of the target variable and try to forecast what will the value be 32 hours after the last measurement.
The second approach is based in the knowledge that each time a measurement of the HB or HM occurs, the process of electrolysis resets. Therefore, every 32 hours 4 measurements of the features are available. Therefore, themodels take as input the last 4 measurements of the features and the previous value of the target variable and try to forecast what the value of the target variable will be after 8 hours.
The third modelling approach is the same as the previous one with the difference that the process resets after each measurement of the (HB or HM) is ignored and more than 32 hours in the past are used.

4.1.1.3.4 Clustering

In the Grant Agreement it is stated that ''*The pots involved in aluminum production line are like individuals behaving according to common rules and trends and individual behaviours, which can drift with time and events, or as a reaction to potline process setting changes.*''. Based on the above statement, makes straightforward to think that the best solution should be obtained after it is possible to find clusters of Pots that have common behaviours.
The techniques utilized to find clusters between the Pots were the K-means, the Dynamic time warping, the Autoencoder and the variational Autoencoder. The best clustering technique found was the DTW (dynamic time warping) and the algorithm was able to extract 6 clusters. For each cluster, a different model was trained. In the following figure, an approximate visual representation of the density of the clusters formed based on ISO map can be observed. It can be seen that there exist twoclusters (black and blue) that are separated enough from the other clusters and contain a low number of pots. These clusters represent pots that are attuned, and these can be treated as outliers.

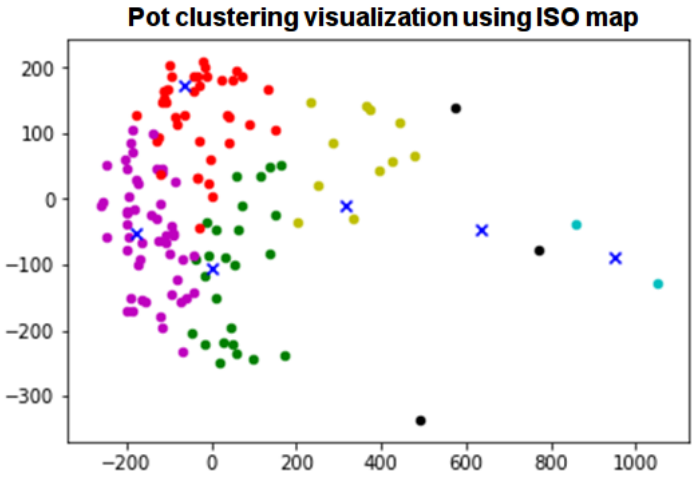

Figure 12 - Clustering visualization for Dynamic Time Warping

The remaining fourclusters are not separated enough in order to observe distinct behaviour of the pots. This could be explained if we regard that, each pot is composed as a combination of behaviours (every possible behaviour) with different weight for each behaviour with the highest weight representing the dominant pot behaviour. Therefore, a pot might fit in cluster of certainbehaviour more than another cluster, but also fit good enough to other clusters and hence the visual representation of the ISO map can be seen inFigure 12. Lastly, the dominant behaviour changes in time and the clustering approach must be adapted on this scenario and monitor when the behaviour of a pot changes in time.

4.1.1.3.5 Models used

The models that were evaluated for the univariate approach are the LSTM, Encoder-Decoder LSTM, RNN, Bidirectional RNN, Dilated Causal Convolutional Neural Network, ARIMA, Linear Regression, Random Forest, Neural Networks and SVM.
The models that were evaluated for both multivariate approaches are the LSTM, RNN, Bidirectional RNN, Encoder-Decoder LSTM, Linear Regression, Random Forest, Neural Networks, SVM'S, Elastic Net and CNN.

4.1.1.3.6 Results

The best results were achieved by using all the features that were available with the second modelling approach (first multivariate approach) without clustering. The reason that using one model for all the pots shows better performance can be explained by two reasons: The first reason is that the clustering method has not reached a sufficient level of classification performance. Another reason that contributes in poorer performance can be explained if it is taken into account that in machine learning theory the larger the number of samples is available the better the generalizations can be achieved. When clustering is performed, each model for each cluster has 3 to 4 times less available data to be trained and therefore it is harder for the model to learn the underlying relations. In the following table, the root mean squared error, the r squared and the mean deviation error metrics for the three best models can be seen. The values of the metrics are the average values after 20 runs of the experiment for each model. Root mean squared error is the mean error between the predicted value and the real value. The R squared statistical metric represents the proportion of the variance in the dependent variable that is predictable from the independent variable. Mean deviation error represents how much the error deviates averagely from the root mean squared error.

Bath height [HBMES]:
Forecasting Bath height is the hardest task among the HBMES, HM and TB variables. It can be seen that results are poor and the models that were developed cannot be used in order to control the Bath Height.

Table 5shows that the best performance is achieved by Artificial Neural Networks with a low mean deviation error relative to the other models.Figure 13, Figure 14 and Figure 15 depict the predicted values (red) and the true values (green) for 200 time steps for the training data (left) and the testing data (right) for each model accordingly. From the figures, it can beobserved that the models struggle to follow the true values outside the range of [14, 20], especially Random Forest that its predictions lie within the range of [15,19]. Therefore, it could be said that the models under-predict outside a certain range of values. Finally, it isobserved that all the models are lacking on following the trend of the time-series.

Table 5 - Bath Height Results

Model name	RMSE Train	R2 Train	MDE Train	RMSE Test	R2 Test	MDE test
Neural Networks (NNR)	1.57	0.54	0.87	1.83	0.37	1.10
RNN	1.83	0.37	1.70	1.94	0.30	1.82
Random Forest (RF)	1.73	0.44	1.92	1.93	0.30	1.85

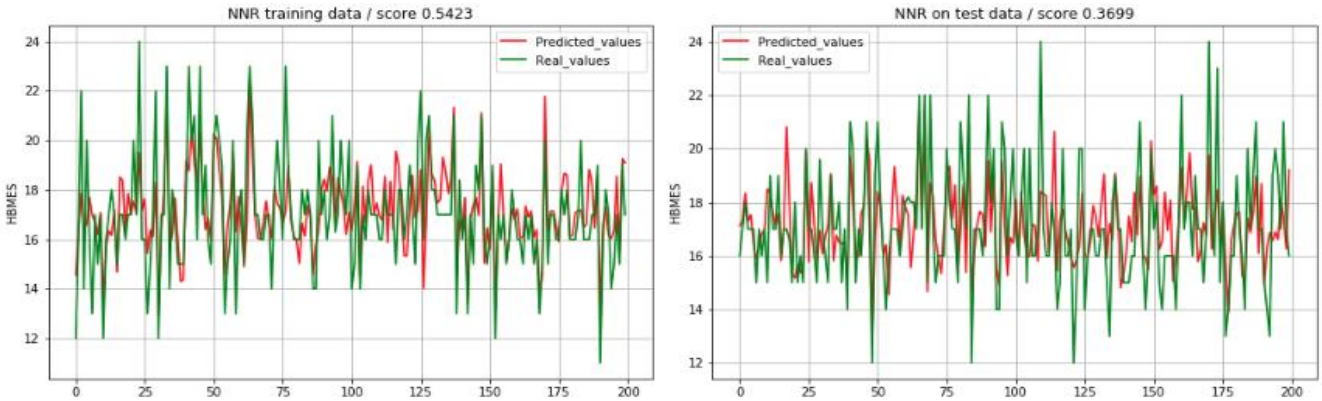

Figure 13 - Artificial Neural Network results for Bath Height. (a) On the Training data, (b) on the Test Data.

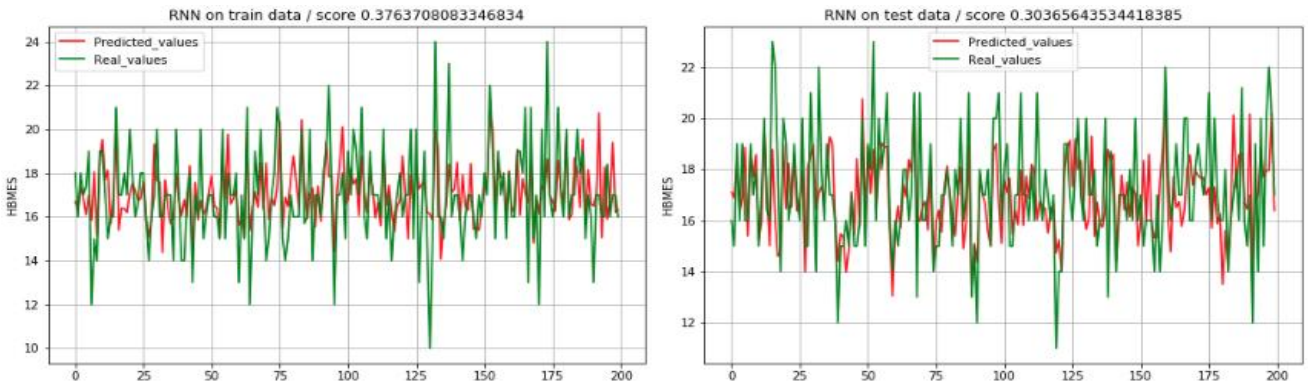

Figure 14 - Recurrent Neural Network results for Bath Height. (a) On the Training data, (b) on the Test data.

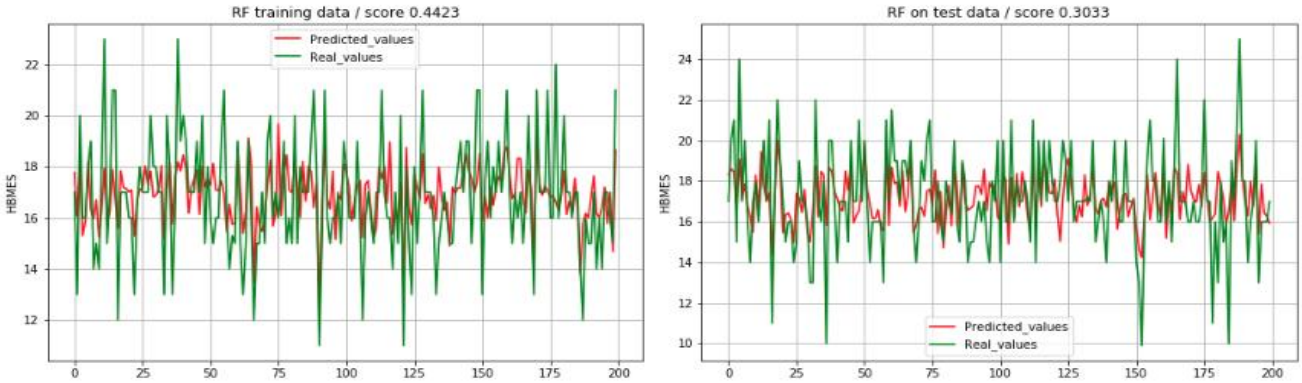

Figure 15 - Random Forest results for Bath Height. (a) On the Training data, (b) on the Test data.

Metal height [HM]:

Metal Height on the other hand can be predicted with a much greater accuracy compared to the other two target variables [HBMES, TB]. Table 6 shows that Artificial Neural Networks also achieve the best performance. As far as concerns the mean deviation error for the RNN and NNR, it can be observed that both models have similar values, but the Random Forest has a much greater deviation relative to the other models.Figure 16, Figure 17 and Figure 18 depict the predicted values (red) and the true values (green) for 200 time steps for the training data (left) and the testing data (right) for each model accordingly. From the figures, it isobserved that the Artificial Neural Networks and the RNN are able to follow the amplitude of the true values but are unable to follow the spikes (values that are too low or too high than the average amplitudes). Random Forest on the other hand is able in this scenario to follow the amplitudes but shows a consistency in overpredicting (predicts values much higher than the true value). Finally, it can be seen thatall the models have a much greater ability in capturing the trend for the Metal Height compared to the Bath Height.

Table 6 - Metal Height Results

Model name	RMSE Train	R2 Train	MDE Train	RMSE Test	R2 Test	MDE test
Neural Networks (NNR)	0.92	0.75	0.71	0.95	0.68	0.83
RNN	1.01	0.70	1.03	1.06	0.67	1.05
Random Forest (RF)	1.09	0.65	1.82	1.93	0.64	1.86

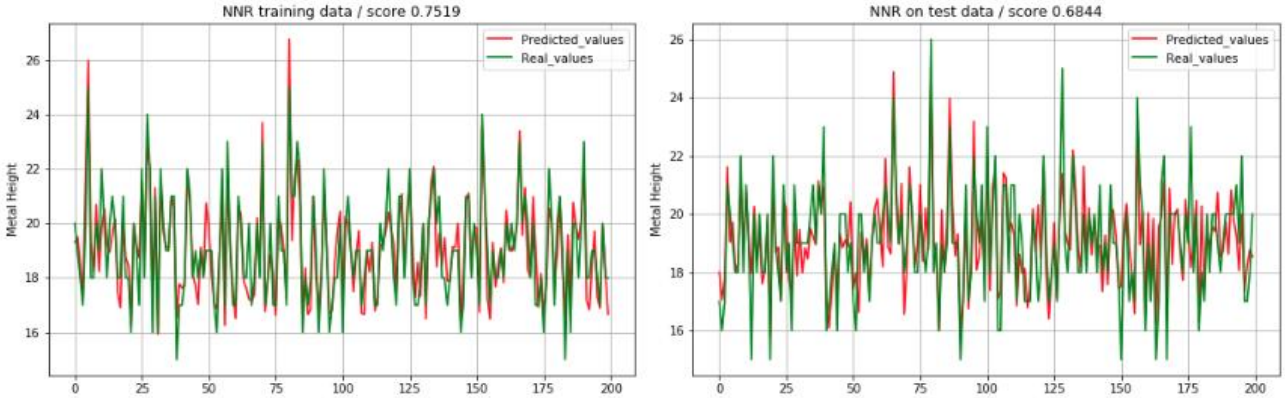

Figure 16 - Artificial Neural Network results for Metal Height. (a) On the Training data, (b) on the Test Data.

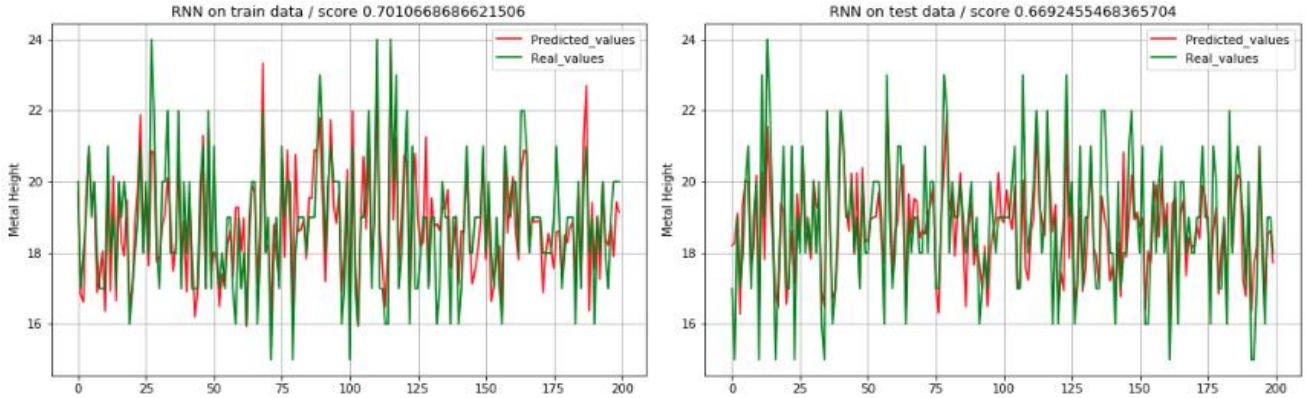

Figure 17 - Recurrent Neural Network results for Metal Height. (a) On the Training data, (b) on the Test data.

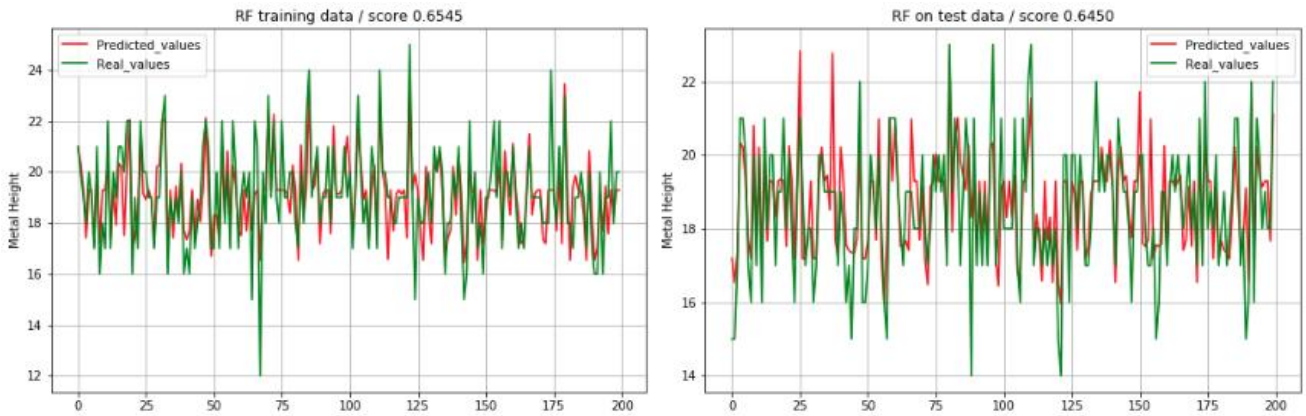

Figure 18 - Random Forest results for Metal Height. (a) On the Training data, (b) on the Test data.

Thermal Balance [TB]:

Finally, for the Thermal Balance there are slightly better results than the Bath Height. The value of the Thermal Balance can be forecasted averagely with +/- 8 degrees deviation from the real temperature, which is not an acceptable range, especially if the mean deviation error is taken into account. Table 7 the results about RMSE, R2 and MDE metrics can be seen. The Artificial Neural Networks also achieve the best performance. This time the second-best model is not an RNN but a Convolutional Neural Network (CNN).

The underlying reason that the CNN as well as RF outperformed RNN is under investigation and it could give a hint towards on what direction to follow to improve our results. The difference of the mean deviation error between the models is proportional to the performance of each model. In contrary, it can be observed that for Metal Height and Bath height scenarios the difference of the mean deviation error does not follow the performance of the models. For example, in the Metal Height scenario Random Forest has almost two times higher value for the mean deviation error than the RNN. Figure 19, Figure 20 and Figure 21 depict the predicted values (red) and the true values (green) for 200 time steps for the training data (left) and the testing data (right) for each model accordingly. From the figures, it can beobserved that no model can follow the true values under the 955 degrees. This is an indicator that a second model should be used to capture the lower temperatures. It can also be observed that CNN is biased with a positive value, where this bias is not seen on the other models.

Table 7 - Thermal Balance Results

Model name	RMSE Train	R2 Train	MDE Train	RMSE Test	R2 Test	MDE test
Neural Networks (NNR)	6.87	0.58	4.37	7.99	0.44	4.83
CNN	7.83	0.46	5.21	8.24	0.39	5.42
Random Forest (RF)	8.68	0.33	5.89	8.83	0.31	6.28

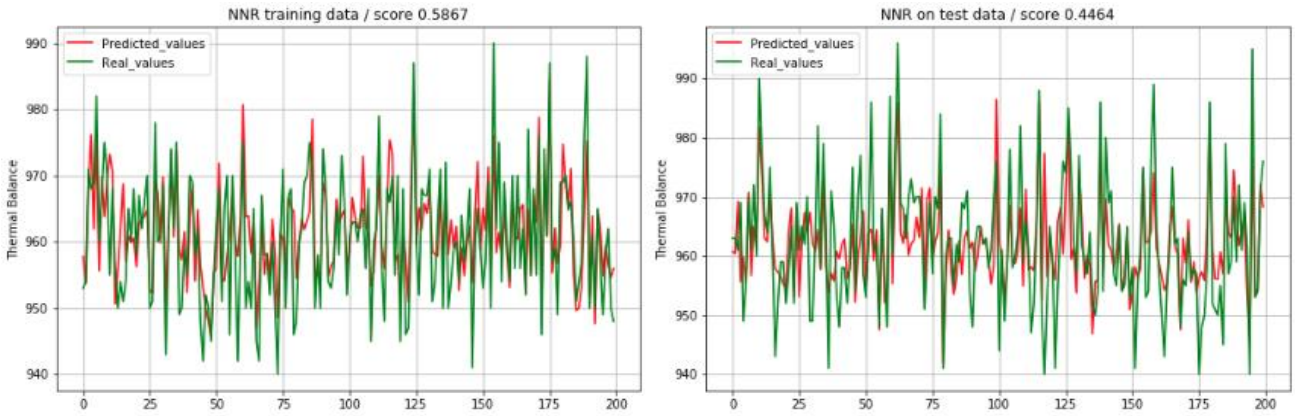

Figure 19 - Artificial Neural Network results for Thermal Balance. (a) On the Training data, (b) on the Test Data.

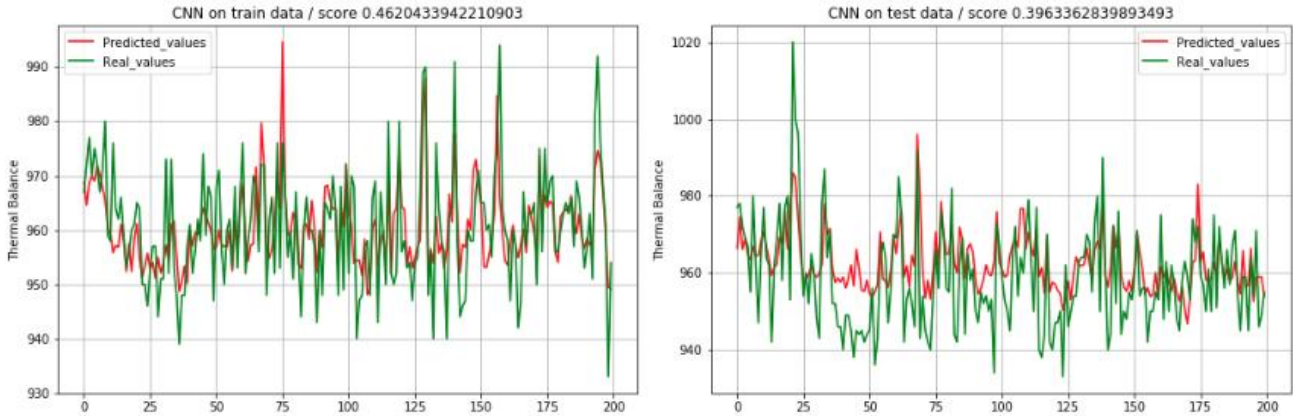

Figure 20 - Convolutional Neural Network results for Thermal Balance. (a) On the Training data, (b) on the Test data.

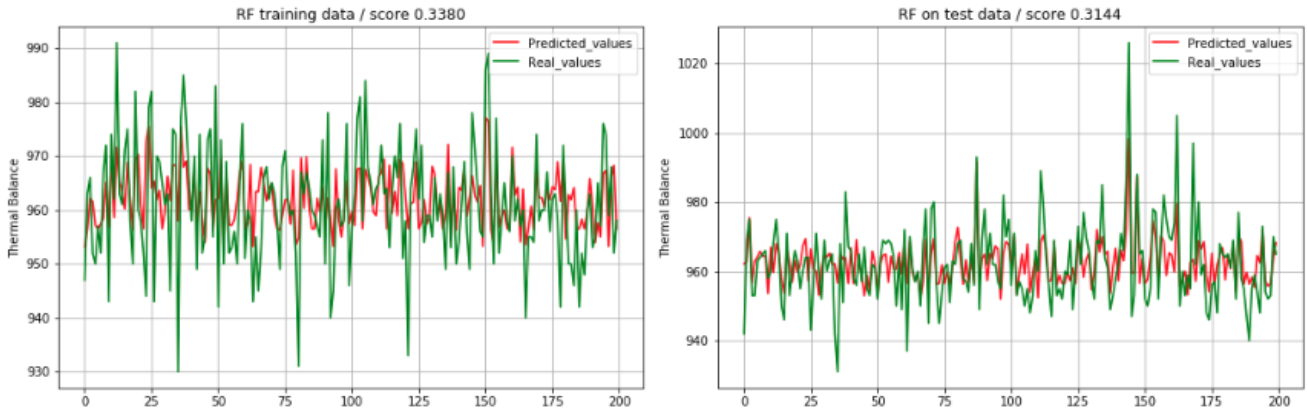

Figure 21 - Random Forest results for Thermal balance. (a) On the Training data, (b) on the Test data.

4.1.1.3.7 Conclusions

A number of preprocessing techniques were performed and the results showed that the models perform when every feature available is kept. Afterwards the univariate and the multivariate modelling approaches of the problem were presented. The best results were achieved with the multivariate approach that takes into account only the last 4 measurements of the features. Consequently, the clustering approach were discussed and it wasfound out that Dynamic Time Warping was the best clustering technique, but the models

performed better without clustering. Finally, the used models were referenced for each modelling approach and the results of the three best models were discussed. In each scenario (Bath height, Metal height and Thermal balance) the Artificial Neural Networks outperformed every other model. The results about Bath Height and Thermal Balance were poor and this indicates that much work has to be done in feature engineering with strong collaboration with the electrolysis experts. The results about Metal Height are encouraging relatively to the other target variables (HBMES, TB) and they can be improved by fine-tuning techniques on the existing model.

4.1.1.3.8 Future Work

- Thermal Balance should be kept ideally within the range of [950 – 970] degrees. When the temperature of a pot exceeds this range, the process of electrolysis is altered and needs some time until it is stabilized again. Therefore, this behaviour should be considered in theanalysis tofind out how long it takes for the transient effect of the Thermal Balance to end and if a single model is able to handle the period of the perturbation of the process.
- For the Thermal Balance and Bath Height variables, the results are poor and therefore new modelling approaches with close collaboration with the electrolysis process experts must be examined.
- New features must be extracted based only on the target variables such as (autocorrelation, absolute energy, first derivative etc.) in order to improve the clustering approach, but also to improve the performance of the process. New features must also be extracted that will be a combination of the current features and will capture a deeper meaning about the evolution of the process.

4.1.2 Anode quality (POC3) and Life Cycle Assessment

4.1.2.1 Anode quality - Approach and results obtained during first iteration

4.1.2.1.1 Approach

For this use case, the aim is to use heterogeneous information from Paste Plant sensors to apply a machine learning methodology, in order to identify and understand root causes of abnormal anode density during the anode production process. By modelling the existing historical data, it is theaim toconsolide the understanding of process experts on anode quality variability, and at finding possible hidden insights.

The approach conducted is a supervised machine learning task, where a classifier can be trained to recognize the quality of anodes produced during a fixed period of time (30 minutes, see below) using the measured sensors signals for that period. The explanation of the possible causes of non-optimal quality is the next step, i.e. the obtaining of an interpretation of the model's behaviour is needed, in order to help the process experts' decisions.

For this analysis, a set of 51 key variables was selected, including for instance BUSS mixer intensities, paste temperatures, and materials flow rates. A data history of 10 months was used, starting from September 2016 and ending on June 2017. The data are grouped in periods of 30 minutes, on which the number of low-density anodes are counted. These periods are the elementary units of the modelling.

As for the POC 2, each sensor's time series on the 30-minutes periods by four very simple statistical features is summarized:

- the median value,
- the standard deviation,
- the 5% and 95% percentiles.

Periods containing paste plant stops are discarded from the analysis: thefocus here is only on normal periods of anode production.

A 30-minute period is considered of "low quality" if at least one anode produced during that period has a green density below 1.620. With this annotation, the following repartition of 'low' and 'good' quality periods can be achieved (Table 8):

Table 8 - Repartition of good and low quality periods in the dataset of POC 3

	Number of periods	Percentage
Low quality	538	9%
Good quality	5423	91%

The classifier selected for this analysis is XGBoost's Gradient Boosted Decision Trees [**Error! Reference source not found.**]. It builds an ensemble of weak classifiers – here, decision trees – in an iterative way. This ensemble model allows to get more robust and more precise results compared to a unique decision tree, and to mitigate the class imbalance problem present in the dataset. The output of the model is a probability of being a "low quality" period.

4.1.2.1.2 LCEN

Optimization of process performance is expected to lower resource consumption in terms of material and energy. To proper quantify these savings, a Life Cycle Assessment (LCA) approach has been adopted and a dedicated life cycle plugin has been added to the MONSOON platform.

The LCA plugin takes as input all the material and energy flows related to the production of the green anode, as well as the output in terms of good and bad quality anodes and related scrap parts. By applying a series of algorithms whose description is out of the scope of this document, the plugin computes a series of environmental KPIs. These indicators are provided per each shift and are visualized in a specific LCA dashboard implemented in Grafana. For the whole description of this component, please refer to [RD.1].

4.1.2.1.3 Performances of the algorithms

The evaluation of the performances of the model was done using a 3-fold cross-validation strategy. The objective is to assess how the performances of the model will generalize to an independent and unseen data set. The original dataset is randomly partitioned in 3 equal size folds containing the same proportions of low/good periods. One partition is retained as a validation set and the remaining 2 are used as a training set. The cross-validation process is then repeated 3 times. As a result, all the periods will be used for training and validation.

It was chosen to look at the precision and recall on the minority class (e.g., "low" quality periods) as performance scores. They are defined as follows:

$$\text{Precision} = TP / (TP + FP)$$

Recall = TP / (TP + FN)

with TP = True Positives, e.g. number of "low quality" periods correctly tagged as "low quality" by the model; FP = False Positives, e.g. number of "good quality" periods tagged as "low quality" by the model; and FN = False Negatives, e.g. number of "low quality" periods tagged as "good quality" by the model. The precision is therefore the fraction of periods predicted as "low quality" by the model that are actually of "low quality"; the recall is the fraction of actual "low quality" periods, among all present in the data, which are retrieved by the model.

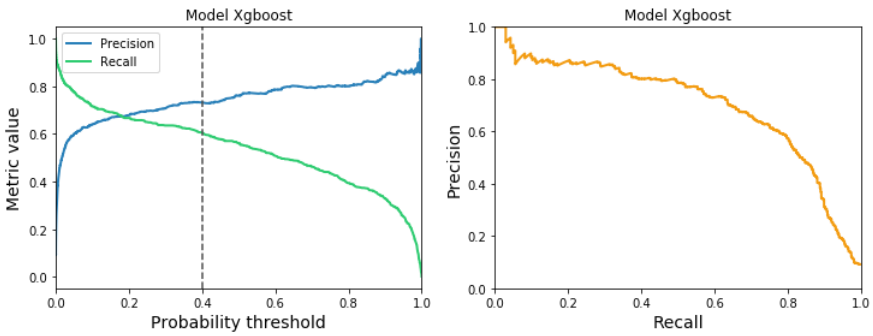

Figure 22 - Precision and recall as a function of the model's probability threshold (left panel); precision as a function of the recall (right panel). Those scores are obtained after a 3-fold cross-validation.

Figure 22 illustrates the performances of the model. The left panel shows the behaviors of the precision and recall, as a function of the probability threshold above which a period is attributed to the class "low quality". As expected, the higher the threshold, the more precise is the model, at the cost of a decreasing recall. The right panel shows the precision as a function of the recall, when varying the probability threshold.

With a threshold of 0.4, the following scores can be obtained:

Precision = 73%

Recall = 60%

Those performances indicate that the model allows finding a significant signal in the data. Therefore,the training of a model has been possible to recognize the process conditions that lead to lower anode quality. It can be used now to detect in a real-time fashion the process deviations, in order to trigger actions in a faster way.

There are few standard parameters to evaluate the general performance of an LCA model. A good rule for performance checking is to track the number of relevant flows covered by the model. For anode quality use case, the LCA is designed to cover most of the relevant[6] material and energy flows in the investigated system, so it is representative for the production of green anode. A task of WP5 activities was dedicated to test the calculations, providing they are in line with robust benchmarks provided by literature, sector studies, manufacturing/business association of aluminium producers. Further description of these activities and of the results obtained is available in [RD.1] and [RD.2].

[6] *Relevant* shall be intended as "major contributor to the majority of most robust environmental KPIs" in the investigated set

4.1.2.1.4 Assessment of the impacts on the process

Once the model trained, the obtaining of the causes is needed that explain the non-quality of the periods in a real-time fashion, for allowing the monitoring of the anode production process at AD. To do so, a module was created, based on LIME (Local Interpretable Model-Agnostic Explanations) [**Error! Reference source not found.**]. LIME allows obtaining interpretations of a model's behaviour by learning an interpretable (i.e. linear) approximated model locally around the instance of interest. For each selected instance, LIME randomly samples instances from the training dataset and weights them according to their distance to the selected instance, and use them for training a locally faithful model. It then deduces from this model some explanations of the model's predicted probability, i.e. a list of variables and associated causes. For example, an explanation produced by LIME could look like this:

- INT-MOY-MALAXEUR–quantile-95 > X
- VIT-MOT-VIS-DEMANDEE–standard-deviation > Y
- DEB-INSTANTANE-DOSEUR–median < Z

This list of features and causes would be sorted by decreasing importance of contribution in the model's decision.

In practice, when the anode quality classifier detects a period of low quality, it triggers the interpretation module, to give in a real-time fashion the causes of low density.

A first version of the function is currently being deployed in the Runtime Container installed at AD. The function is automatically running every thirty minutes. A user interface for visualizing the results of the function was developed, using Grafana [**Error! Reference source not found.**], an open-source platform for creating analytics dashboards.

The LCA plugin computes a set of environmental KPIs. Even though the whole list is available in deliverables D 5.7 and D 7.1 ([RD.1] - [RD.2]), an overview of most robust and relevant indicators is reported below. Numeric results are computed referring to one kg of good anode produced (functional unit)

Table 9 - Most relevant environmental KPIs for anode production

KPI	Unit	Description
Global Warming Potential	mass of equivalent CO_2	Total amount of equivalent greenhouse gases emitted to produce 1 kg of good anode
Acidification Potential	mass of equivalent SO_2	Potential of acidification of the air arising from the production of 1 kg of good anode
Primary Energy Demand	equivalent MJ	Overall amount of primary energy required to produce 1 kg of good anode

This approach is expected to provide valuable insights for process experts and environmental managers about environmental effects of process optimization. Results are visualized in a specific Grafana dashboard which main screen is reported below (Figure 3 and 4).

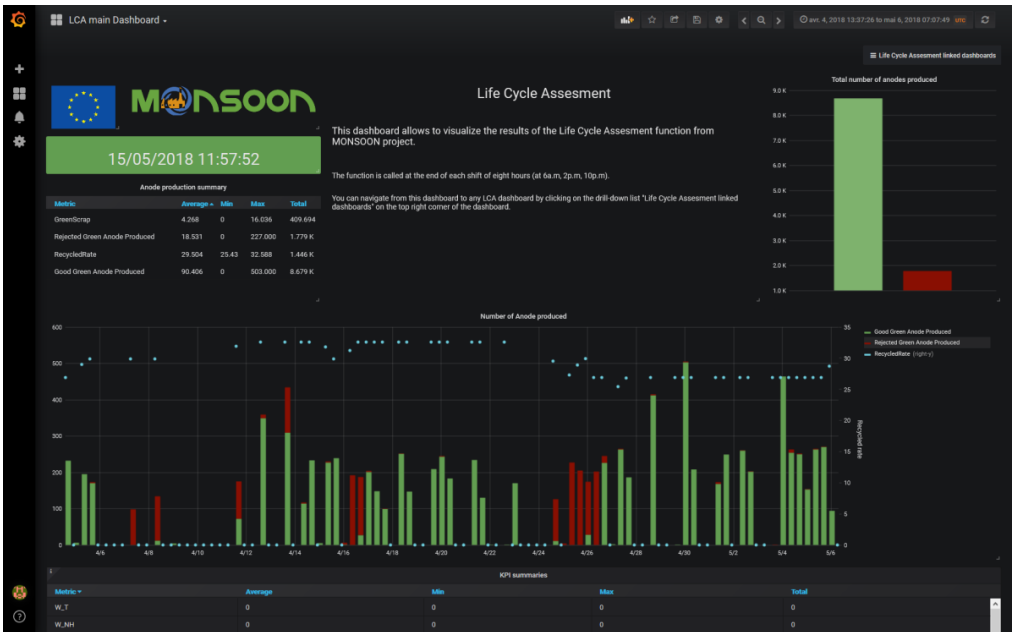

Figure 23 - Main LCA dashboard for anode production

In this visualization, bottom bar plot shows the shift in a selected time period; green bars are related to good anodes, while red ones are about bad anodes that will be scrapped and recycled in the production loop. On the top section, left box shows some process-specific KPIs such as gross production of good and bad anodes, as well as recycling rate. Each KPI is reported with its minimum, maximum and average value in the investigated time. Right box gives the overview of production of good and bad anodes in time.

By selecting a specific KPI, the dashboard automatically switches on this visualization:

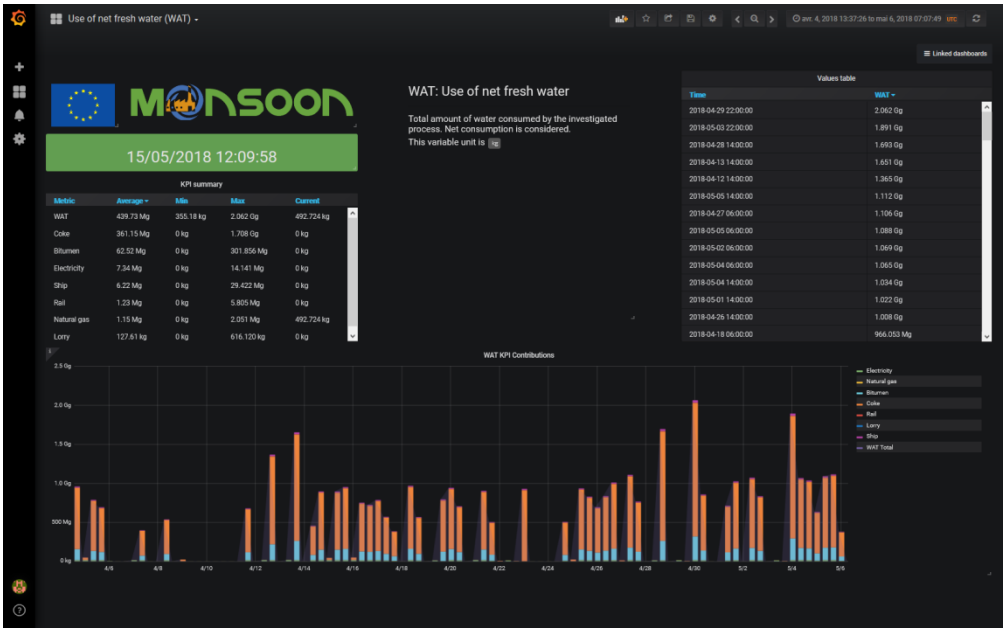

Figure 24 - KPI-specific dashboard for anode production

This dashboard shows specific results of the selected KPI. In the left, upper part an overview of records is reported with a breakdown of relevant contributors (electricity, coke, pitch, transport, gas, etc.). This breakdown is reported in the bottom part with shift by shift visualization. Right, upper box shows the numeric results for each shift for the selected KPI.

By navigating through these dashboards, plant users and managers can visualize trends of environmental figures; these insights can be used to support sustainability practices or to demonstrate the environmental compatibility of optimization strategies proposed by data scientists.

4.1.2.2 Anode quality – on-going developments for second iteration

The current version of the predictive function for anode quality allows process experts to get a first understanding of the process parameters that deviate from nominal conditions, impacting anode density, in an online manner. However, going from those "explanations" to concrete corrective actions is a challenging task: most of the parameters used for the modelling of anode quality are "sensors", which cannot be directly modified by operators. When a period of lower anode density appears, the ultimate goal would be to propose a list of **actionable** parameters to modify, and how, in order to increase the density for the next periods.

To do so, a new model for anode density has been developed. Instead of predicting if a 30-minutes period is of "low" or "high" anode quality, thedirect prediction the median density of the anodes being produced is possible. The problem is now a regression task. The process parameters used as input variables are the same as for the previous classifier, but only the median values are computed to summarize their behaviours in each period – this is an easier statistical feature for giving recommendation.

Among all the input process parameters, Aluminium Dunkerque defined the ones which are actionable, as well as the ranges on which they can act.

The approach being investigated so far is presented in Figure 25. For a given period of low anode density, for each actionable parameter A, the range of values accessible for it is scanned, as defined by Aluminium Dunkerque. Furthermore, all the sensors are varied which are strongly correlated to A, in the vicinity of their values for the period. For each scanned value, the regression model is applied to get the estimate of the density associated to this combination of parameters.

Then the three actionable parameters that impact the most the density for this periodare determined. Finally,the re-simulation of the model predictions is performed for all the combinations of those three parameters and select the combination that leads to the highest anode density.

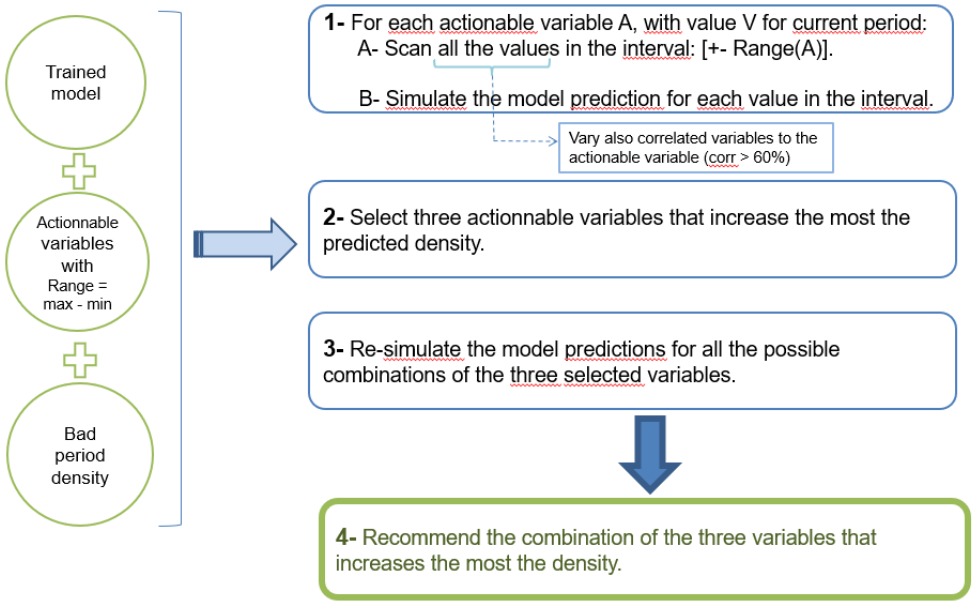

Figure 25 - Approach for anode density optimization

This approach has been tested on the 680 periods with median density < 1.625 in the September 2016 – June 2017 dataset. According to the regression model, the obtained recommendations could lead to an increase of density of approximatively 0.007.

These preliminary results could be improved by investigating the following points:

- Taking into account interactions between actionable parameters for selecting the top 3
- Implementing constraints on the actionable parameters
- Adding the full list of actionable parameters defined by Aluminium Dunkerque to the model
- Add information about individual anodes raw material using MESAL system

Moreover, after finding during POC 2 analysis (see **Error! Reference source not found.**) that some process parameters can change strongly as a function of time, it is necessary to be able to automatically re-train the models for anode density. On this important task during iteration 2 will be worked.

Anode process has been fully modelled with LCA approach during these months of the MONSOON project. Some enhancements of the component may occur during incoming months; these updates could follow feedback received by Aluminium Dunkerque users about additional KPIs that might be tracked to increase the value of the tool for environmental management. Most demanding obstacle to face is that connection with MONSOON and sensors collecting information must be ensured to compute additional KPIs (such as dust emissions); even though this is not a blocking issue, this may require some effort by infrastructure partners that was not foreseen at the beginning of the project.

4.2 Plastics domain

4.2.1 Coffee capsules

In this section the different approaches followed to understand the production process of coffee capsules and the lids shall be discussed. There are multiple sources of coffee capsule data. Two sources are the readings injection molding machines, namely HUSKY and kraussMaffei. Furthermore, the quality information of each capsule and lid is captured in the IMD vista machines.

Though there are no much concrete results with the data analysis in plastic domain, the efforts are worth recording.

4.2.1.1 Results obtained

4.2.1.1.1 Trend analysis

Trend analysis includes methods to find interesting patterns and features from the raw or processed data from the plastic production machines. Here some of the trend analysis are discussed which were applied on the plastic data.

Repeating patterns

Since the data from the plastic moulding machines do not come labelled, it is important to find some interesting repeating patterns in the data so that it can be shown to plastic domain experts for further investigation. The implementation of STOMP [1] algorithm finds out the top matching and repeating patterns in a univariate time series. The time series data from plastic domain is used to find the interesting patterns. Below figure shows an example where the stoppages are identified in the attribute "*plastification time*" of KraussMaffei machine. Here a unique pattern is automatically identified as a top motif and similar patterns can are deducted with the similarity search algorithms.

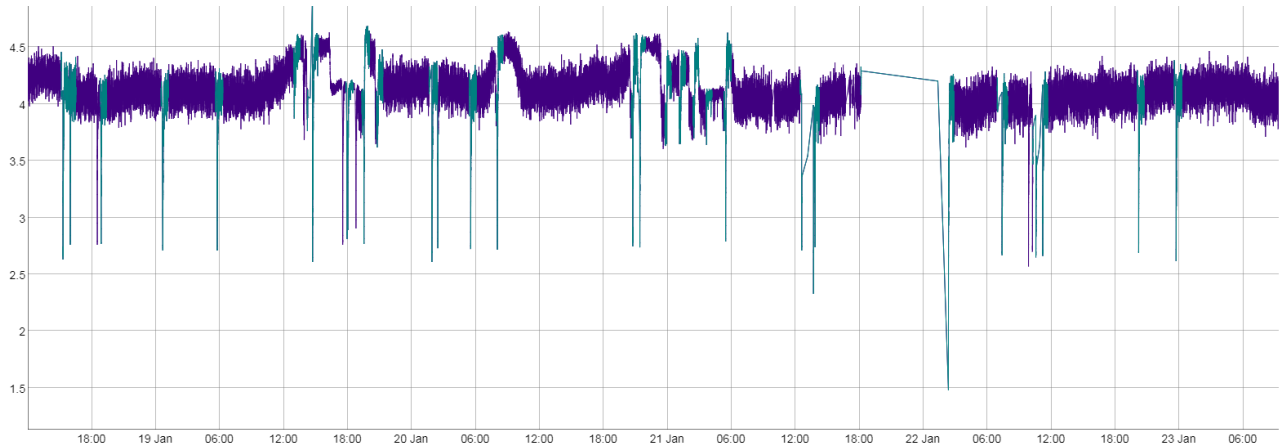

Figure 26 - Plastification time highlighted with possible stoppages in the machines. The Y axis shows the plastification time in seconds.

4.2.1.1.1.1 Exploratory analysis

Different statistical methods are applied to the raw data from injection moulding machines to identify interesting patterns. These patterns are then shown to the domain experts for their opinions. During this process, unlabelled data is visualized in different ways and shown to the domain experts so that they get

more insights. Although this has been done in locally by data scientists, this shall be done in future in the data lab. It would be easier for the data scientists to share their exploration and visualization results to the domain experts and analyse it in a collaborative way. Different tools such as Grafana and Zeppelin notebooks facilitate this approach.

4.2.1.1.2 Modelling the Plastic Production Process

Multiple attempts were made to modelling the plastic production process. The attempts here are described.

4.2.1.1.2.1 Prediction of Time-To-Live based on Cycle Interruption messages in HUSKY machines

Idea is to define segments of HUSKY machine data based on the cycle interruption messages. Cycle interruption messages are considered to be raised whenever the machine stops due to multiple reasons (either manually or automatically stopped). The entries within segments are ignored if they have any gaps. The TTL (Time-To-Live) for each segment using the upcoming (nearest) cycle interruption is calculated. Afterwards a RNN based on the machine data is trained with TTL value as the target.

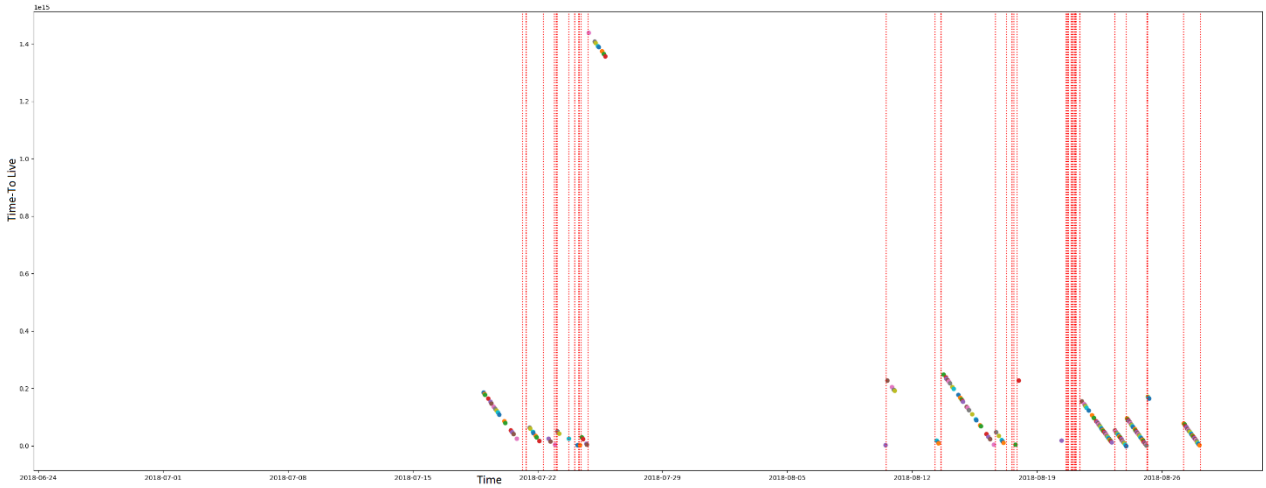

Figure 27 - Labelling of the data with time to live before a cycle interruption. The red vertical dotted lines represent the cycle interruption and the dots represent the time to live for each cycle.

Cycle interruptions are represented by vertical dotted red lines. Gaps between segments are due to missing values or noise. One can see the TTL value for each segment (as large integer values according to the y-axis), where the TTL of a segment is proportional to the distance to the next cycle interruption. So far, the trained model could not find any correlation between the machine data and cycle interruptions. Ruben confirmed this result, since most of the cycle interruptions were performed by manual checks/fixes of the operators and can therefore not be identified using the data. With this itcan be concluded that the cycle interruptions are not correlated with machine observations.

Nonetheless, further attempts are made by data scientists for finding out the relation between different events recorded in HUSKY machines to the recorded events. So far there are no positive results where there is a high correlation between the sensor readings and the machine observations.

4.2.1.1.3 Correlation between Injection moulding machine data and IMD Vista Data

The logs from IMD vista stating their quality is pre-processed for a binary classification problem, wherein if the amount of invalid products is beyond a threshold for a specific period. Whenever there is an interval with a high rejection rate according to IMDVista, the associated machine data is classified as invalid and vice versa. Different statistical features are extracted from the two classes(mean, standard-deviation etc.). Features are then compared against each other to check for dissimilar representative machine observations regarding the two states (valid, invalid). In addition, segments of both classes were inspected manually in order to potentially identify a distinct machine behaviour based on the provided machine data.

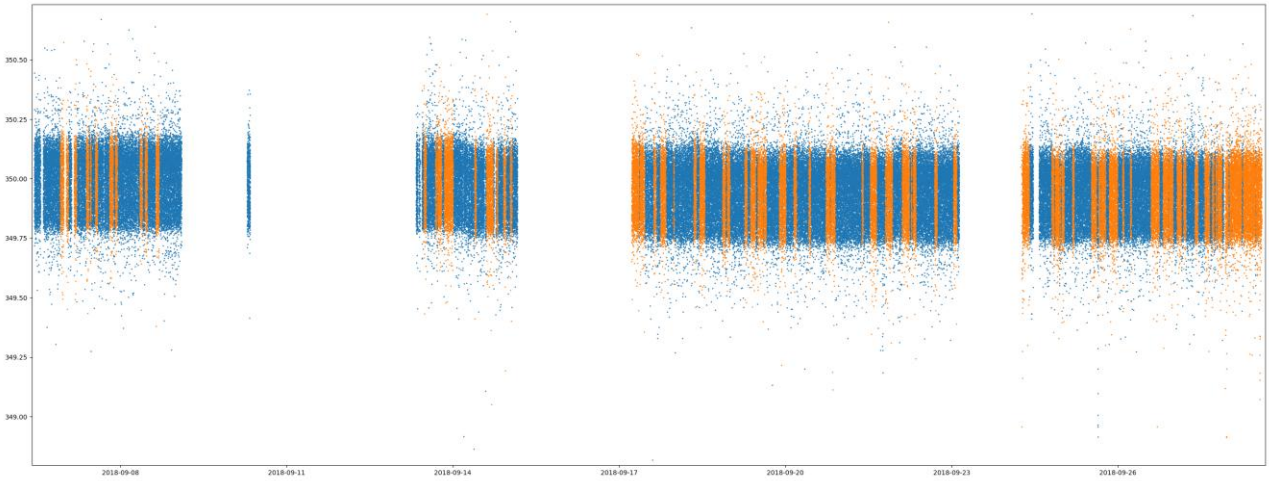

Figure 28 - Labelling one of the HUSKY machine parameter to valid(orange) and invalid (blue) segments based on the number of quality capsules produced

As of now, correlation between machine data and high rejection intervals of the IMDVista system could be not be identified. The feature representatives are very similar for valid and invalid. This also holds for improved tolerance values and a domain-based matching of machine data dimensions to specific quality parameters.

4.2.1.2 Improvements

As explained in previous sections, no concrete results were obtained to model the Coffee capsule and lid production process. The investigations are still going on in the direction of finding the configurations corresponding to the lowest defects. This can only be possible if the relation between Injection machine readings and the quality of the products during these readings are found. Some of the machine learning methodologies which find the non-linear relations between the features and outputs need high amounts of input data. This has been a challenge since the quality data is recorded in the later stage of the project. As more quality information is available, the models can predict the relations with better accuracy. Once the modelling of the manufacturing process is successful, this can be used in alerting the floor operators about possible raise in defects in future.

Other possible improvements include the early detection of anomalies. This shall be done by fitting a model for the injection moulding process and forecast the measurements for a future time. If the forecasted measurements go beyond threshold, alerts can be generated. This is one of the future works.

5 Training aspects

5.1 Introduction

Training is an important part in order to facilitate the adoption of a change. However, in our case since the Monsoon project is an ongoing project, new features are added regularly. It is a hard task then to determinate what should be the final content of the training modules before all features are available.At this step of the project all the final features are not added yet.
In the following paragraphs the training planwill be detailed as discussed with the Monsoon's consortium in October 2018. The approach chosen by the consortium is to first identify the target group which will be interested by the Monsoon platform and then try to address their specific training needs.

5.2 Target group definition

The identified targets for training sessions are:

- **ESG members:** this group is composed of industrial partners selected in the beginning of the project to be part of the Monsoon project. For those members the training is more a demo phase of the platform in order to generate a desire to use the Monsoon platform. This can be done by organising physical sessions to present the Monsoon platform and demonstrate its performances.

- **End users:**those persons are the final users of the Monsoon platform. Theirs technical needs are bigger than the ESG members and the training modules must be technical and more specific. This group has to be divided in subgroups to address theirs needs and to perform a better training.

Aluminium Pechiney and GLN experts identified three end-users' subgroups, based on their own organisation charts and knowledge of the running of their own industry. Those three groups are:

- **IT team**:the IT teamshave a crucial role in the industrial environment: they mission in to ensure the maintenance of the production network. The integration of the Monsoon platform into a production network directly depends on their abilities to administrate the platform. Those users have big technical training needs and the training modules must be as compete as possible.
- **Process team** are the ones that will be using the Monsoon platform in order to visualize the predictive functions, analyse their data with the support of the visualisation tool: creating new dashboards, create alerts, in fact use the visualisation tool to maximise the control on processes… For them the administration of the platform is out of the scope, and they should be trained to use the Monsoon platform in the most efficient way.
- **Environment team**: their role is to follow the evolution of the gas emissions, raw materials waste and all the possible impact of their process from an environmental point of view. They are linked to the corresponding legal authorities. Here the Monsoon platform offers them a monitoring of their main KPIs thanks to the LCA plugin. In that scope they have to be able to use the visualisation tool and to understand the construction and meaning of the LCA plugin.

5.3 Training plan

Once those groups have been identified specific uses cases need to be addressed for those groups. With this vision it is possible to link the training modules and the different use cases, both for Plastic and Aluminium domain. The next step is to identify training modules in order to have one training material per module. With those modules the fulfilling of the user needs, and the buildingofa full training plan can be started. And finally train the specific userswith their specific modules.

However, those steps are not done yet at this state of the project. They will be described in the D7.4 Final demonstrator for Aluminium and Plastic domains deliverable, at the end of the project.

6 Conclusions

This deliverable documents in detail the integration of the Monsoon platform into two industrials environment. Indeed, after detailing the proof of concept for each use case and for each domain was done through the engagement charters.

Thanks to technical hypothesis the assessing of the impact of the models and predictive functions developed by Monsoon on both plastic and aluminium domain through the Proof of Value is reached.

Based on this deliverable an industrial validation, both for Plastic and Aluminium domain, that the solutions provided by the Monsoon platform are viable and can be interesting for future customers is available.

Now the improvement of the performance and the relevance of the models and develop now POC for the second iteration of the project is possible.

Those improvements will be highlighted in the D7.4 deliverable: Final Demonstrators in the Aluminium and Plastic domains.

Acronyms

Acronym	Explanation
POC	Proof of Concept
POV	Proof of Value
HSE	Health, Safety & Environment
OPEX	Operational Expenses
CAPEX	Capital Expenditure
KPI	Key Performance Indicators
LCA	Life Cycle Assessment
SVM	Support-Vector Machine
HB	Bath Height
HBMES	Measured Bath Height
RF	Random Forest
CNN	Cellular Neural Network
RNN	Recurrent Neural Network
SVM	Support Vector Machine
TTL	Time To Live

List of figures

List of tables

References

[1] Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, 785–794 (2016) http://doi.acm.org/10.1145/2939672.2939785

[2] Ribeiro, M.T, Singh, S. and Guestrin, C.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, 1135–1144 (2016)

[3] Grafana website: https://grafana.com/grafana

Appendices

List of variables to be used as input data for the POC2 (and possibly the POC1) of the aluminium domain (+ corresponding thresholds):

1. Buss mixer
   - D110-J160_INT_MOY_MALAXEUR (POC 3)
     - Target: 620A
     - No min and max values for the moment. To be defined by the data analysis.

   - D110-J160_INT_MOY_MALAXEUR (POC1 or 2)
     - Target: To be defined with data analysis
     - Plugging tendency warning: >380A
     - Plugging emergency alert: >500A.

   - D110-H080_TEMP_F_T_MALAXEUR (POC1 & 2)
     - Target: 190C°
     - Low temperature default: <140C°.
     - High temperature default: >245C°→ pump stops.

   - D110-G130_TEMP_BRAI_CORIOLIS_G130 et D110-G131_TEMP_BRAI_CORIOLIS_G131 (POC3)
     - Target: 185°C
     - Low temperature threshold: <140°C (not currently followed in AD).
     - High temperature threshold: >245°C (not currently followed in AD).

2. Eirich cooler
   - D110-J170_NIVEAU_MELANGEUR_REFROIDIS (POC 1 or 2)
     - Target: adjusted automatically with the tool intensity (next parameter)
     - Set point threshold: 5-10% under the high-level value set up by the operator (max value possible: 85%).
     - Cooler high-level default: > set up value by operator during 20s (max operator set up value 75%). → BUSS and dry product stop.

   - D110-J172_INT_MOT_TOURBILLON (POC3)
     - Target: 185A.
     - Low intensity threshold:<183A
     - High intensity threshold:> 187A
     - High intensity default: >250A during 30s (on "aggregated" data over 10s). → BUSS and dry product stop.

   - D110-J170_TEMP_INSTANT_PATE (as a replacement for D110-J173_DEB_EAU_REFROIDISSEMENT_J173) (POC3)
     - Target: 156°C.
     - Low paste temperature default - 5C° (during 1min).
     - High paste temperature default + 5C° (during 1min).
     - High temperature alarm + 10C° above set point.

3. Vibrocompactors

   - D110-K060_VIT_VIBRATION et D110-K070_VIT_VIBRATION (POC3)
     - Different values of target for each vibrocompactors (may evolve a little regularly)

- Target K060: 1335tr/min,
- Target K070: 1320tr/min.

- D110-K030_POIDS_NET_TREMIE (POC3)
  - Target: 1075kg
  - Low paste weight: - 10kg
  - High paste weight: +10kg