H2020 – ICT-13-2018-2019

MUSKETEER

Machine Learning to Augment Shared Knowledge in Federated Privacy-Preserving Scenarios (MUSKETEER)
Grant No 824988

D7.5 Use case execution and KPI evaluation in the Smart Manufacturing domain

November 21

Imprint

Contractual Date of Delivery to the EC: **30 November 2021**

Author(s): **Giacomo Fecondo (FCA), Chiara Napione (COMAU), Lucrezia Morabito (COMAU)**

Participant(s): **Giacomo Fecondo (FCA), Chiara Napione (COMAU), Lucrezia Morabito (COMAU), Susanna Bonura (ENG), Davide Profeta (ENG), Jaime Medina (TREE)**

Reviewer(s): **Ambrish Rawat (IBM), Emil Lupu (IMP)**

Project: **Machine learning to augment shared knowledge in federated privacy-preserving scenarios (MUSKETEER)**

Work package: **WP7**
Dissemination level: **Public**
Version: **1.0**

Contact: **giacomo.fecondo@stellantis.com**
Website: **www.MUSKETEER.eu**

Legal disclaimer

The project Machine Learning to Augment Shared Knowledge in Federated Privacy-Preserving Scenarios (MUSKETEER) has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824988. The sole responsibility for the content of this publication lies with the authors.

Copyright

Executive Summary

The deliverable D7.5 – Use case execution and KPI evaluation in the Smart Manufacturing domain - is the result of tasks T7.3 and T7.4 related to automotive manufacturing domain. These tasks deal respectively with the deployment, at pilot site, of components that enable the MUSKETEER Federated Machine learning platforms services and an overall evaluation of the demonstrations.

It is based on the validation of the requirements driven by the KPI defined checklist. The evaluation has been done following both a technical perspective and a business one.

Starting from the Smart Manufacturing use case description, the procedure to run the pilot with the MUSKETEER platform has been detailed and the results presented. Some end-user feedbacks are provided about the platform usage and the technical results of the federated algorithm.

Document History

Version	Date	Status	Author	Comment
0.1	24/05/2021	First draft ToC	FCA-COMAU	First draft
0.2	15/09/2021	Toc update	ENG-FCA-COMAU	Update
0.3	05/11/2021	Overall check before internal review	FCA-COMAU	Update
0.4	16/11/2021	Feedback from internal reviewers	IMP-IBM	Review
1.0	23/11/2021	Final version	Mark Purcell & Gal Weiss	Final review

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

Abbreviation	Definition
ANN	Artificial Neural Network
API	Application Programming Interface
CA	Consortium Agreement
CC	Client Connector
FML	Federated Machine Learning
GQM	Goal-Question-Metrics Model
IDSA	International Data Spaces Association
KPI	Key Performance indicators
ML	Machine Learning
PERT	Program evaluation and review technique
PK	Public Key
RSW	Resistance Spot Welding
POM	Privacy Operation Mode
SMEG	Smart Manufacturing Evaluation Group
WN	Worker Node

1 Introduction

1.1 Purpose

This document is the description of the results of the execution and the evaluation of project final prototype version in Smart Manufacturing domain.

The pilot evaluation activities have been mainly based on the MUSKETEER evaluation framework, defined in D2.7, built on key technical quality focuses and business priorities identified for accurate evaluation, questions and metrics.

1.2 Related Documents

This deliverable, describing the output of tasks T7.3 and T7.4, takes into account what was presented in the following documents (also see Figure 1) and consolidated during the project:

- D2.1 Industrial and Technical Requirements
- D2.3 Key performance indicators selection and definition
- D2.7 Key performance indicators selection and definition - final version
- D6.2 Scalability of machine learning algorithms over every POMs
- D7.4 Final prototype of the MUSKETEER client

In particular FCA ITEM and COMAU, as end-users, evaluate the project results on the basis of the defined KPIs in terms of measurable metrics.

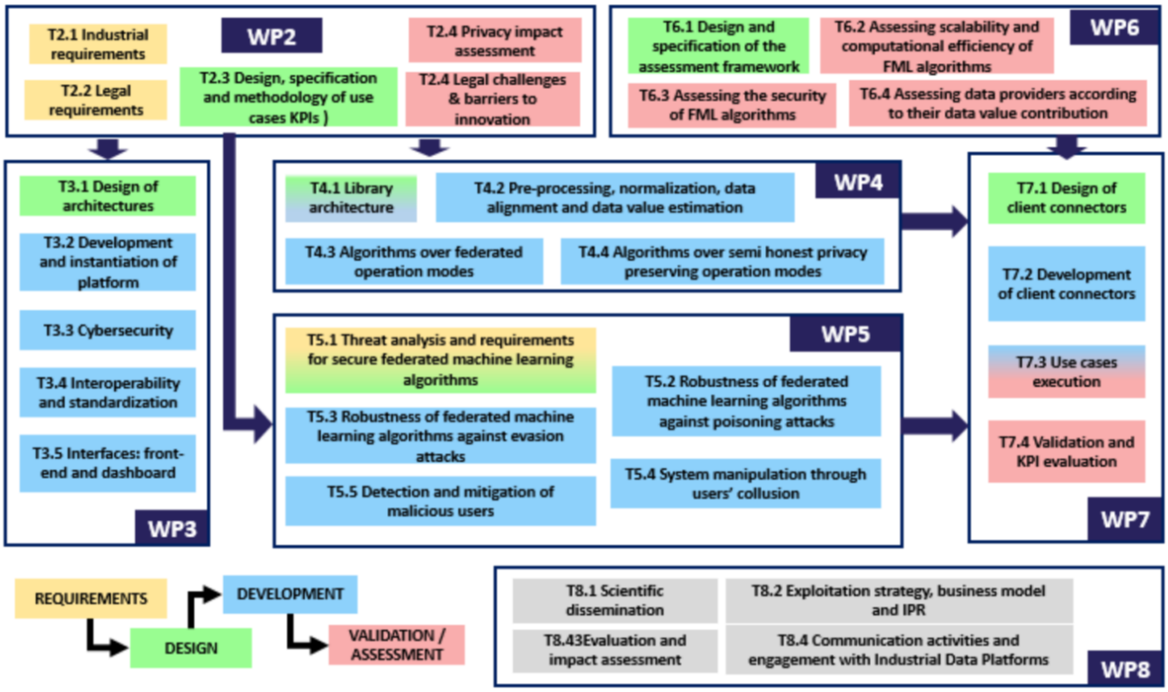


Figure 1 - MUSKETEER's PERT diagram

1.3 Document Structure

The document is organized as follows. In the next section (Section 2) the MUSKETEER evaluation perspectives are presented according to two principal views: the technical view and the business view.

In Section 3, business general performance indicators are recapped and an evaluation is provided by the industrial end-users FCA and COMAU.

In Section 4, a summary of Smart Manufacturing case is described in order to have a landscape picture that enables the comprehension of the industrial domain on the one hand and on the other of how MUSKETEER platform responds to industrial requirements providing ML algorithms and adequate privacy operation modes. The core of the section is the pilot setup and test execution that provide end-user feedback on MUSKETEER platform usage.

Section 5 deals with measurement goals, questions and metrics for the evaluation of MUSKETEER results measured in the Smart Manufacturing use case.

Section 6 presents a description of the technical perspective.

Finally, in Section 7 we provide some conclusions reached during the execution of T7.3 and T7.4.

2 MUSKETEER Evaluation perspectives

The MUSKETEER Evaluation Framework contains four phases inspired by the GQM method (cf. D2.7 for further details):

- Planning phase, executed within WP2.
- Definition phase, executed within WP2.
- Data Collection phase, addressed within WP7 and described in this document.
- Interpretation phase, addressed within WP7 and described in this document.

In accordance with the MUSKETEER Evaluation Framework for validating the MUSKETEER platform, each evaluation objective defined for the smart manufacturing use case will be validated according to two different perspectives that are defined below:

(1) The Technical Perspective, in which some aspects of the MUSKETEER architecture, together with Privacy preserving FML algorithms under different POMs and FML Algorithms will be evaluated.

(2) The Business Perspective, in which the response to the user will be examined. The techniques used in this perspective are drawn upon usefulness and feasibility of the proposed deployment operation support.

Both perspectives will be taken into account in Section 5, while in the following section a set of general and domain-independent performance indicators, initially introduced in D2.7, will be evaluated.

3 Business General Performance Indicators

The following table presents the quantitative and qualitative evaluation metrics that correspond to the evaluation of the MUSKETEER platform operation phase for product validation purposes. In general, many of the general performance indicators that are adopted refer to the ISO/IEC 25010:2011 standard [3] for usability aspects, and are measured in a qualitative manner, either by measuring AS-IS and TO-BE values, or in case of more qualitative answers, by using a 1-5 scale. In the last column, we show the evaluation done by the end-users FCA and COMAU.

Characteristic	Identifier	Metric	Definition	Mandatory (Y/N)	Evaluation (October 2021)
	Business Value				
Clarity	GPI_BusV1	Clarity level	How clear was it for you what the MUSKETEER platform is about? [Scale 1 (Little) -5 (Very)]	Y	**5**
Added Value	GPI_ BusV2	Added value level	How much added value do you feel that the MUSKETEER platform provides to your operations while using it? [Scale 1 (Low) -5 (High)]	Y	**4**
Need Importance Level	GPI_ BusV3	Need importance level	How important is for you the need that the MUSKETEER platform covers for you? [Scale 1 (Little) -5 (Very)]	Y	**5**
Need Coverage	GPI_ BusV4	Need coverage level	To which degree does the MUSKETEER platform cover your needs? [Scale 1 (Low) -5 (High)]	Y	**4**
Innovation	GPI_ BusV5	Innovation level	How innovative do you find the idea of the MUSKETEER platform?	Y	**5**

			[Scale 1 (Little) -5 (Very)]		
Intention to use	GPI_ BusV6	Intention level	To what extent do you intend to use the MUSKETEER platform? [Scale 1 (Low) - 5 (High)]	N	**4**
Virality	GPI_ BusV7	Virality level	How probable is it for you to recommend the MUSKETEER platform to other stakeholders? [Scale 1 (Low) -5 (High)]	Y	**4**
	Effectiveness				
Effectiveness	GPI_Effe1	Effectiveness level	Is the MUSKETEER platform enabling you to accurately achieve your goals for data sharing and data Analytics [Scale 1 (Low) -5 (High)]	Y	**4**
	Efficiency				
Efficiency	GPI_Effi1	Efficiency level	Is the MUSKETEER platform efficiently fulfilling its intended Purpose? [Scale 1 (Low) -5 (High)]	Y	**4**
	Satisfaction				
Usefulness	GPI_Sati1	Usefulness level	Do you find the MUSKETEER platform useful? [Scale 1 (Low) -5 (High)]	Y	**4**
Trust	GPI_Sati2	Trust level	Do you trust the MUSKETEER platform and its provided functionalities? [Scale 1 (Low) -5 (High)]	Y	**5**
Pleasure	GPI_Sati3	Pleasure level	Does the MUSKETEER platform please you when	Y	**3**

			you use it? [Scale 1 (Low) -5 (High)]		
Comfort	GPI_Sati4	Comfort level	Do you feel that the MUSKETEER platform provides a comfortable user interface and workflows? [Scale 1 (Low) -5 (High)]	Y	**3**
	Freedom from risk				
Economic damage risk	GPI_Free1	Level of economic damage risk	How sure are you that MUSKETEER protects you from exposing you on economic damage? [Scale 1 (Low) -5 (High)]	Y	**3**
Privacy harm risk	GPI_Free2	Level of data privacy damage risk	How sure are you that MUSKETEER is protecting your data privacy? [Scale 1 (Low) -5 (High)]	Y	**5**
	Usability				
Learnability	GPI_Usab1	Learnability level	How easy it was for you to learn how to use the MUSKETEER Platform? [Scale 1 (Not) -5 (Very)]	Y	**4**
Flexibility	GPI_Usab2	Flexibility level	How much do you believe the MUSKETEER platform can be used for applications other than the demonstrator ones? [Scale 1 (Low) -5 (High)]	Y	**5**
	Content Conformity				
Quality	GPI_Conte1	Content quality	How useful do you find the data and the applications found in the MUSKETEER platform in terms of quality?	Y	**4**

			[Scale 1 (Little) -5 (Very)]		
Quantity	GPI_Conte2	Content quantity	How satisfied are you by the quantity of the data and the algorithms found in the MUSKETEER platform? [Scale 1 (Little)-5 (Very)]	Y	**3**

Table 1 –Domain independent performance evaluation

4 Smart Manufacturing use case

4.1 The reference scenario

Nowadays, the use of automation, in particular robots, is becoming more and more widespread in the automotive industries. All the advantages that robotics can bring to a factory, like FCA's, derive from the thousands of person-hours needed to configure the machines and to maintain them, avoiding loss in quality. In fact, maintenance is crucial to preserve the production from incredibly huge downtime costs. Unfortunately, maintenance has a high cost. However, it is possible to reduce it, thanks to the help of machine learning techniques. Predictive models, if trained on significant historical data, will be able to alert of a possible future failure or quality decrease, enabling a more efficient maintenance. Since a robot manufacturer, like COMAU, creates instances of the same application in different sites (e.g., robotized welding system), it clearly appears the potential advantages of combining historical data of all these instances, to train one more generic model, able to improve the overall impact on the production quality. All the instances mentioned can come from different plants of the same car manufacturer or from plants of different car manufacturers. This could bring to a benefit for all the companies involved and also to the supplier of the equipment, in this case represented by COMAU. However, the scenario described requires a data collection and analysis platform that is reliable and secure, to guarantee data protection for all the stakeholders. Sharing and analysing data must favour the creation of a reference model, based on AI techniques, which, appropriately fed and trained, through the use of shared data, is able to guarantee the quality for that specific operation, at desired levels and with a positive impact on the final process.

4.2 The welding use case

The purpose of this use case is to collect and analyse the data related to the welding process, to search, with the support of AI technologies, correlations among the variables that characterize the process so that a produced welding point will be of the expected quality.

In particular, the choice of resistance spot welding process derives from the fact that it is an efficient joining process widely used for the realization of sheet metal assemblies. Low cost, high speed and suitability for automation make RSW an attractive solution for auto-body assemblies and several other appliances. The high number of spot welds (3000:7000) in a car body requires in any case that process parameters must be fine-tuned.

Welding is the process by which two pieces of metal can be joined together thanks to a fusion of the layers. A welding gun is composed by:

- two mechanical arms, one fixed and the other which can move;
- a linear motor which allows the arm movement;
- a copper electrode at the end of each arm, which is in contact with the metal sheets to weld;
- a water-cooling system;
- a welding tray which is the controller of the current supplied for the welding.

Figure 2 - Welding gun

In general, the number of metal sheets to weld varies from 2 to 3. The current supplied, the time spent on the welding process and the pressure applied by the arms on the metal sheets strictly depend on the number of layers and on their thickness. The current is supplied by the welding tray and flows through the arms up to the pieces of metal to melt. The spot-welding time cycle is characterized by four time-measurements: squeeze time, weld time, hold time and off time.

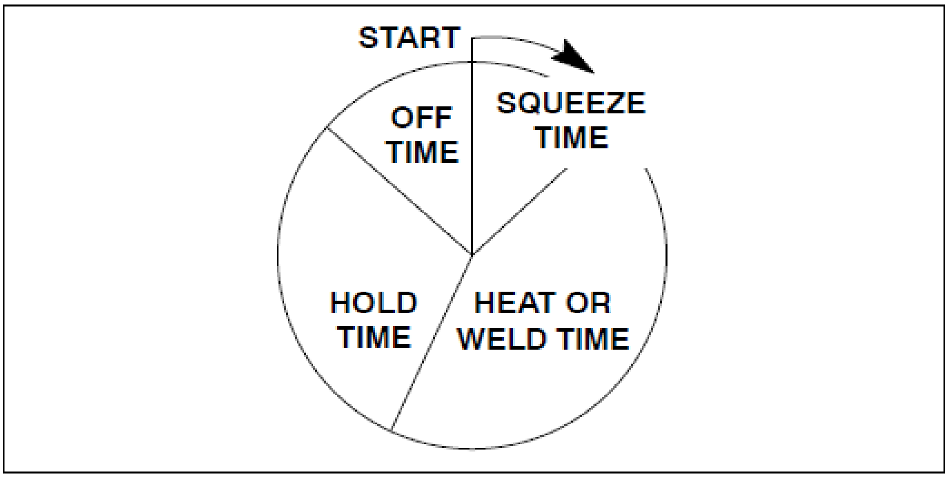


Figure 3 - Spot-welding time cycle

The squeeze time represents the time between the pressure application and the weld; the weld time represents the weld time in cycles; the hold time represents the time when the pressure is maintained after the weld operation is completed and the off time occurs when the electrodes are separated to enable the next spot. During each welding point, the electrodes are subject to degradation and they get dirtier. This causes a loss of quality in the following welding points. For this reason, after a predefined number of points, the electrodes undergo to a dressing process, which consists in the removal of small material. After several removals the electrode has to be changed. Welding data contain information on these processes by means of counter variables. In general, RSW is based on four major factors, which most describe the welding process:

- amount of current that passes through the “work piece” [kA];
- time in which the current flows through the “work piece” [s];
- pressure that electrodes apply on the “work piece” [daN];
- the area of the electrode tip contacts with “work piece”.

Obviously, the aim is to get a quality weld and therefore the key is to configure the parameters so that the target can be achieved: so, the material properties of the sheets must be taken into account (contact resistance and electrical resistivity). The Figures 4, 5 and 6 show the contact resistance of mild steel sheets, stainless steel sheets and aluminium sheets as a function of pressure at different temperatures [4]. Then, the electrical resistivity of the different materials as function of temperatures is shown in Figure 7.

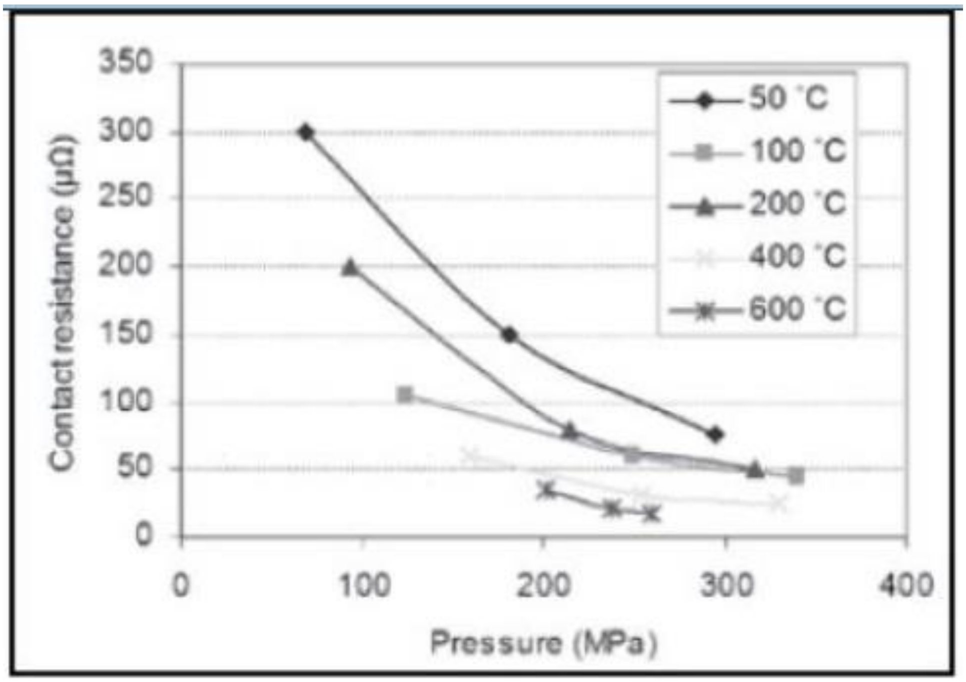


Figure 4 - Contact resistance of mild steel sheets

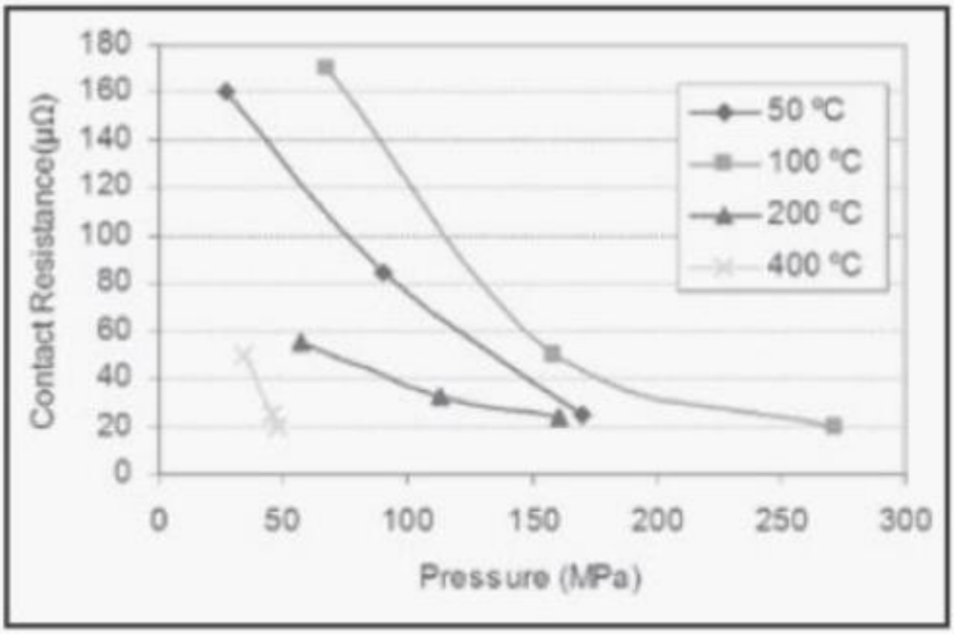


Figure 5 - Contact resistance of stainless-steel sheets

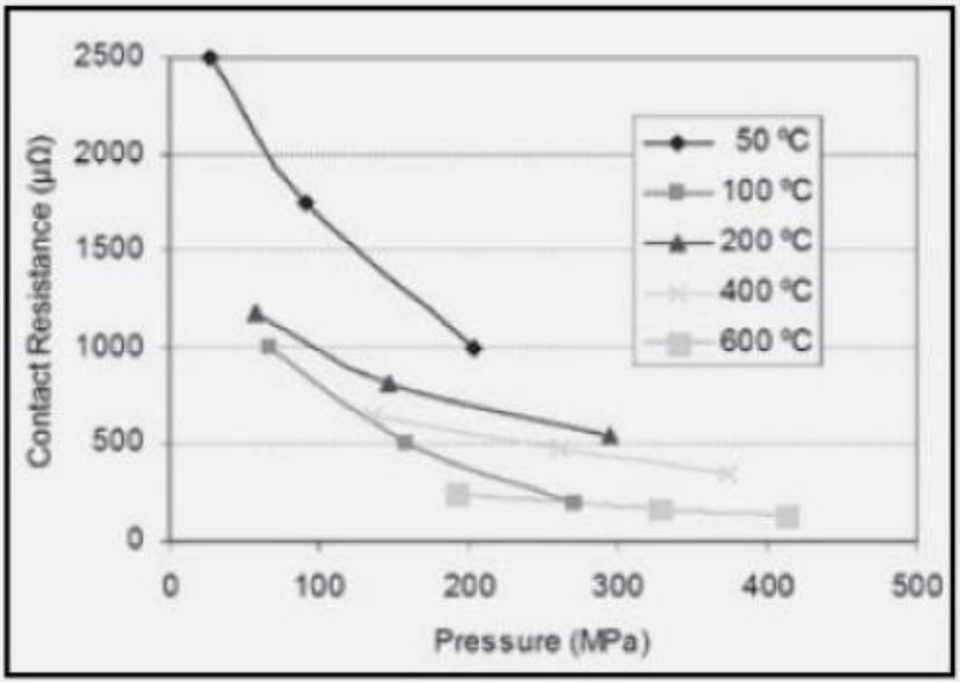


Figure 6 -Contact resistance of aluminium

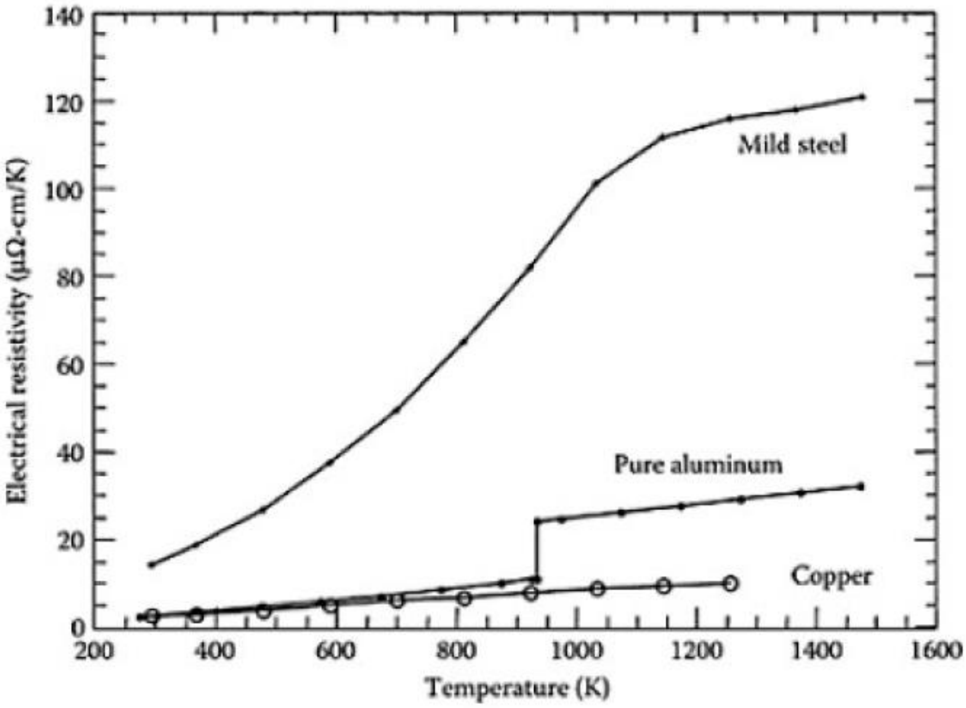


Figure 7 - Sheets resistivity as function of temperature

The phase of setting up of welding guns and of reference "master" curves for a quality weld represent a fundamental activity that must be repeated when the context changes significantly (e.g., sheet thickness).

To give one example, the Figure 8 represents a typical (total) resistance curve of a successful weld. It's taken from the programming manual for the setting of parameters involved.

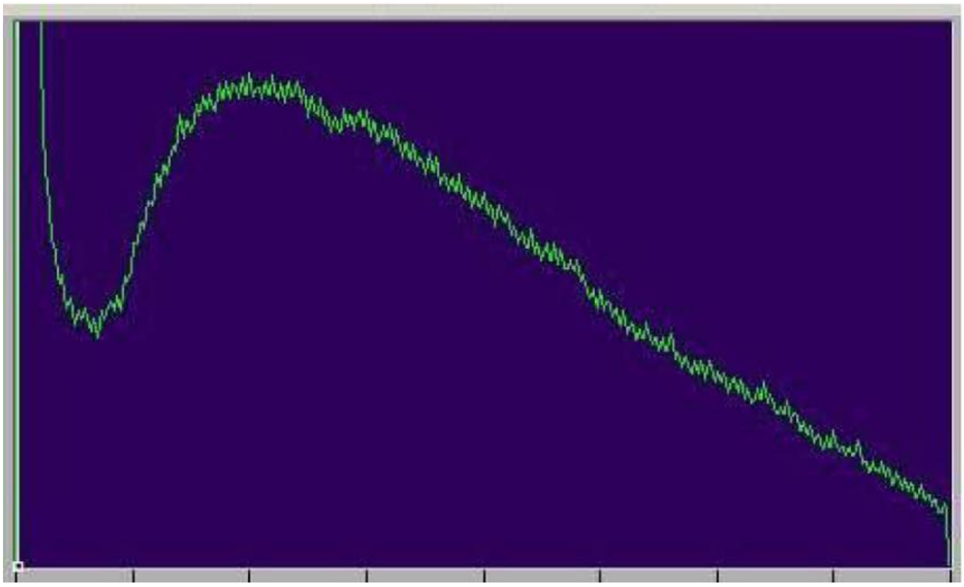


Figure 8 -Resistance curve of a successful weld

Figure 9 shows more details of process events that characterize the behaviour of the curve (y-axis: resistance/x axis: welding time).

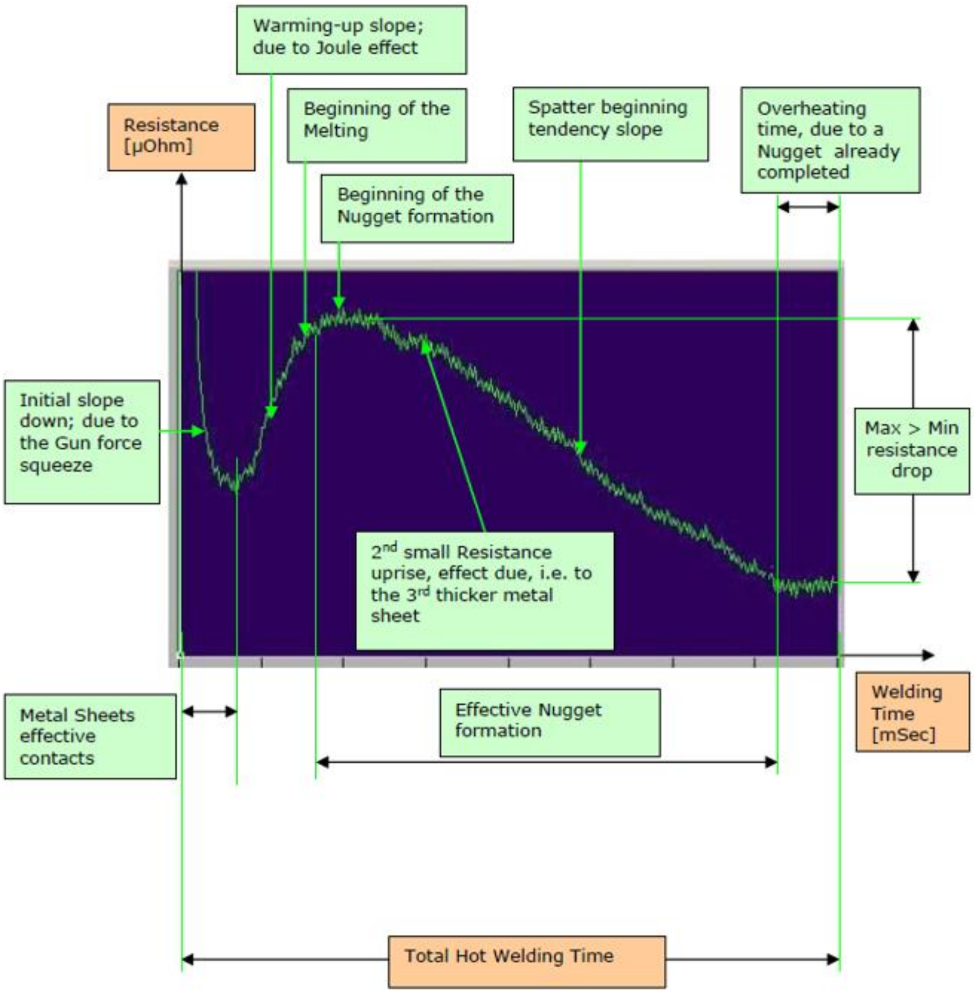


Figure 9 - Resistance curve and welding process events (characteristic points)

4.2.1 Relevance of federated machine learning in welding use case

Having clarified the reference scenario, we focus our attention on the maintenance process that can benefit from the use of machine learning models.

FCA and COMAU share their know-how to identify and develop methods and techniques for data collection and data analysis related to the welding process. These data have to be considered across different plants, in order to highlight, with the support of AI technologies, correlations among the variables that mostly characterize the process. In this way, welding points will reach the expected quality and it will then be possible to estimate the remaining time to loss of quality. The availability of the sought correlations enables the improvement of the welding process with positive impacts both on the quality of the welding process and on the final product associated with it. In particular, the data sharing and analysis could generate an error probability distribution curve, in qualitative terms. The curve is stabilized within a set interval such that the impact on the final product quality is acceptable. Moreover, the estimation of the remaining time to loss of quality can improve the maintenance organization, avoiding replacements of still functioning pieces and, on the other hand, avoiding unexpected failures.

This will have a direct impact in the reduction of the maintenance of the manufacturing process due to:

1. An improvement of the welding process with positive impacts on both the quality of the welding process and on the final product associated with it;
2. A reduction in the welding gun maintenance cost. With regard to privacy and security concerns, since historical records contain information about the industrial processes of a company and the solutions used, any information leakage can potentially reveal industrial secrets about internal manufacturing processes and the problems that production teams have to face in the plants. This information can give a competitive advantage to OEM competitors and cause damages to the brand. This is why we use IDSA concepts and models to ensure confidentiality and privacy protection to the IDSA ecosystem stakeholders. In addition, a factory may be hit by cyber-attacks. A data poisoning attack can produce a useless predictive model and the malfunction of the production plant. An attack can led to a false alarm of a possible future failure and a maintenance cost increasing. On the contrary, it can also lead to predicting that maintenance is not required, when in fact it is, and the consequent failure of the component as well as a lower quality produced output.

4.3 Pre-processing, ML Algorithms and POM

The data used for the manufacturing use case comes from welding systems located in the FCA plant located in the South of Italy. In particular, the welding systems chosen are the ones which weld the right and the left side of a vehicle. These two systems do exactly the same operation. For each side, the chosen robot welds 9 points. The starting dataset consists of all the data extrapolated, cycle by cycle, from XML files in which the chronological sequence of the operations, carried out during the welding process, is recorded.

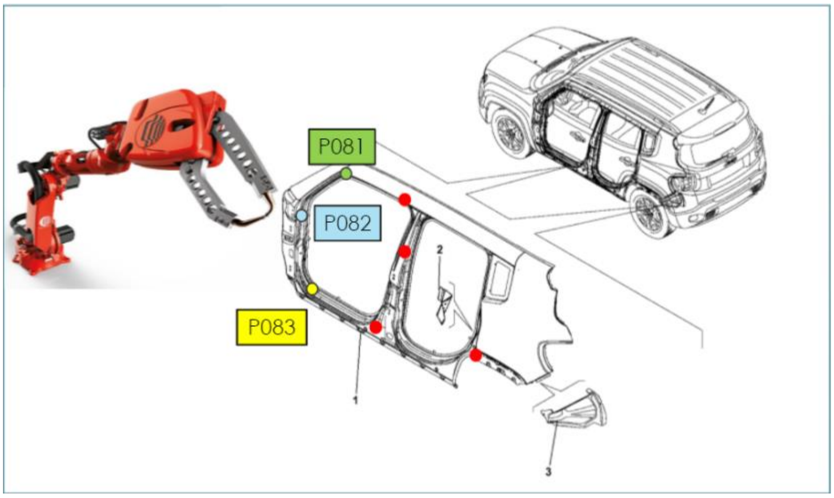


Figure 10 – Use case representation: welding robot and the body side with the highlighted welding points.

Each file contains the following variables as features:

Variable ID	Variable description	Variable type	Measure unit
DEPCOD	Department code	String	-
SPOTNUM	Welding point code number (univocal for each point of the cycle)	String	-
PRGNUM	Welding program number	Int	-
WELTIME	Welding point timestamp	Timestamp	-
WFORC1	Set value of the electrodes force	Int	DaN
WFORC2	Set value of the electrodes force	Int	DaN
REGMOD	Check it quality regulator mode (on-off)	Bool	-
CLSMOD	Check it quality classifier mode (on-off)	Bool	-
ALARMCODE	Alarm code	Int	-
WELMOD	Program execution mode (Normal, Monitor, Corr. Cost)	String	-
CURMOD	Check current mode (Switch off, primary current, secondary current)	String	-
WELCNT	Welding points counter	Int	-
WDRSCNT	Welding dressing counter	Int	-
WELR	Resistance	Int	uOHM
TOLDINF	Diagnostic lower tolerance percentage of the current	Int	%

TOLDSUP	Diagnostic lower tolerance percentage of the current	Int	%
WELTIME1_S	Settled welding time phase 1	Int	ms
WELTIME1_M	Measured welding time phase 1	Int	ms
WELCUR1_S	Settled welding current phase 1	Int	kA
WELTIME2_S	Settled welding time phase 2	Int	ms
WELTIME2_M	Measured welding time phase 2	Int	ms
WELCUR2_S	Settled welding current phase 2	Int	kA
WELTIME3_S	Settled welding time phase 3	Int	ms
WELTIME3_M	Measured welding time phase 3	Int	ms
WELCUR3_S	Settled welding current phase 3	Int	kA
WELTIME4_S	Settled welding time phase 4	Int	ms
WELTIME4_M	Measured welding time phase 4	Int	ms
WELCUR4_S	Settled welding current phase 4	Int	kA

As target variable or label the following variable is used:

Variable ID	Variable description	Variable type	Measure unit
WCLSRES	Quality Index	Int	-

The pre-processing phase of the data includes the following steps:

- Read all XML files to create two unique CSV file, one related to right welding robot and one related to left. Only the following variables, a subset of the original ones, are selected to be used, as features, in the algorithm training phase.

Variable ID	Variable description	Variable type	Measure unit
ID_POINT	Welding point code number, previously called SPOTNUM	String	-
WELCNT	Welding points counter	Int	-
WDRSCNT	Welding dressing counter	Int	-
WELCUR1_S	Current	Int	kA

WELR	Resistance	Int	uOHM

- Discretize the Quality Index value based on the classification done by the welding system supplier on the outcome of the welding spot: spatter welding due to a high thermal energy compared to that of the measured in the setup phase; good welding if it has an acceptable quality; stuck welding due to a low thermal energy compared to that of learning phase; cool welding due to short circuit welding.

We show below the different types of welding quality and their relationship to the quality index values.

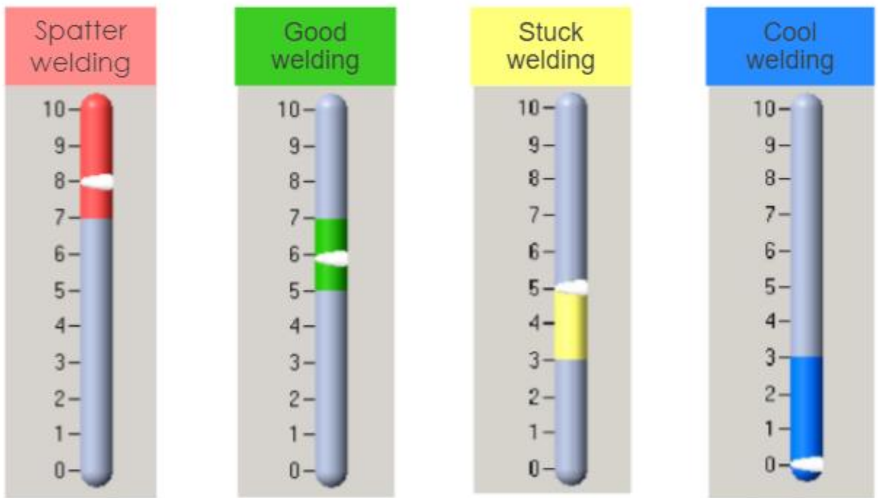


Figure 11 – Quality Index

```
dataD.loc[dataD['WCLSRES'] <= 5, 'label'] = 0

dataD.loc[(dataD['WCLSRES'] > 5) & (dataD['WCLSRES'] <= 7), 'label'] = 1

dataD.loc[dataD['WCLSRES'] > 7, 'label'] = 2
```

However, due to the lack of data with Quality Index < 3, we have only used 3 labels.

- Apply to the target column a one hot encoding transformation.

After these transformations, the dataset appears as follow:

ID_POINT	WELCNT	WDRSCNT	WELCUR1_S	WELR	label_0	label_1	label_2

P001	87	73	9.36	329.8	0	1	0
P003	88	73	9.46	323.8	0	1	0
P005	89	76	9.27	331.2	0	1	0
…	…	…	…	…	…	…	…

The last phase of data preparation consists in splitting the datasets into training (~75%), validation (~10%) and test (~15%) set, balancing the data based on the label columns.

The smart manufacturing use case aims to predict the quality of the welding point so the class of algorithms needed is a classification one. In particular, a three layers Neural Network has been chosen.

The NN adopted in the final evaluation of the platform has the following shape:

- (None, 48); activation function: "relu"
- (None, 16); activation function: "relu"
- (None, 3); activation function: "softmax"

Finally, the most suitable POM, taking into account the confidentiality requirements of organizations involved (FCA and COMAU), at this stage of knowledge, is PORTHOS (POM3).

In fact, this POM is able to protect the final model of each participant, that has to be private. Moreover, we hypothesize that the workers don't trust each other (because they could potentially belong to different legal entities or different companies), so the encryption has to be implemented with different private key for each data owner (cf. D4.1- 5.1.3).

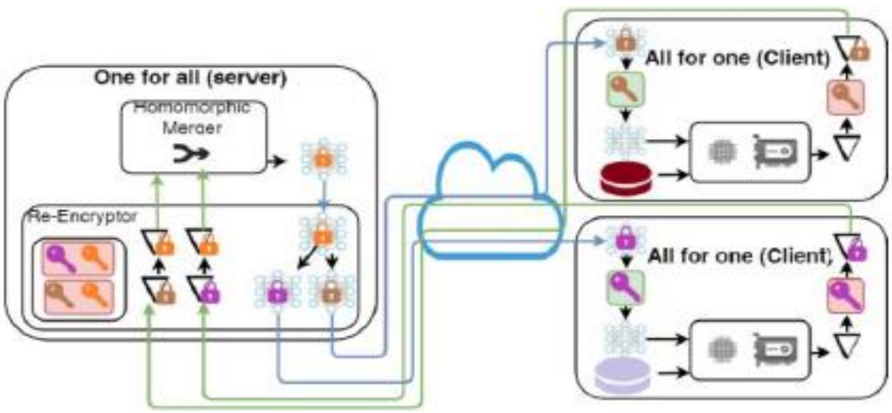


Figure 12 - Porthos, the selected privacy operation mode

4.4 Data set

To summarize, the raw data consist on XML files as follow:

Origin	Number of XML files	Dimension of single file	Time Horizon
Left side welding robot	49627	≈ 15.5 KB	From November 2019 to February 2020
Right side welding robot	30983	≈ 15.5 KB	From September 2019 to February 2020

The final data files obtained are the following:

Filename	File format	Dimension	Dataset shape
aggregator_test_idpnt.csv	csv	3.2 MB	(99814,8)
aggregator_val_idpnt.csv	csv	2.9 MB	(89451,8)
S_worker_train_idpnt.csv	csv	7.3 MB	(229065,8)
S_worker_val_idpnt.csv	csv	896 KB	(27920,8)
S_worker_test_idpnt.csv	csv	1.2 MB	(37227,8)
D_worker_train_idpnt.csv	csv	2.9 MB	(102363,8)
D_worker_val_idpnt.csv	csv	364 KB	(12917,8)
D_worker_test_idpnt.csv	csv	485 KB	(17223,8)

4.5 Pilot setup and execution

In this section the pilot setup and execution of the smart manufacturing use case are explained in detail, for platform validation.

Figure 13 shows the platform architecture with all its components and how the communication during a task is managed. In the context of this specific deliverable, we are mainly focusing on the Client Connector node: in fact, the end users only interface with this tool.

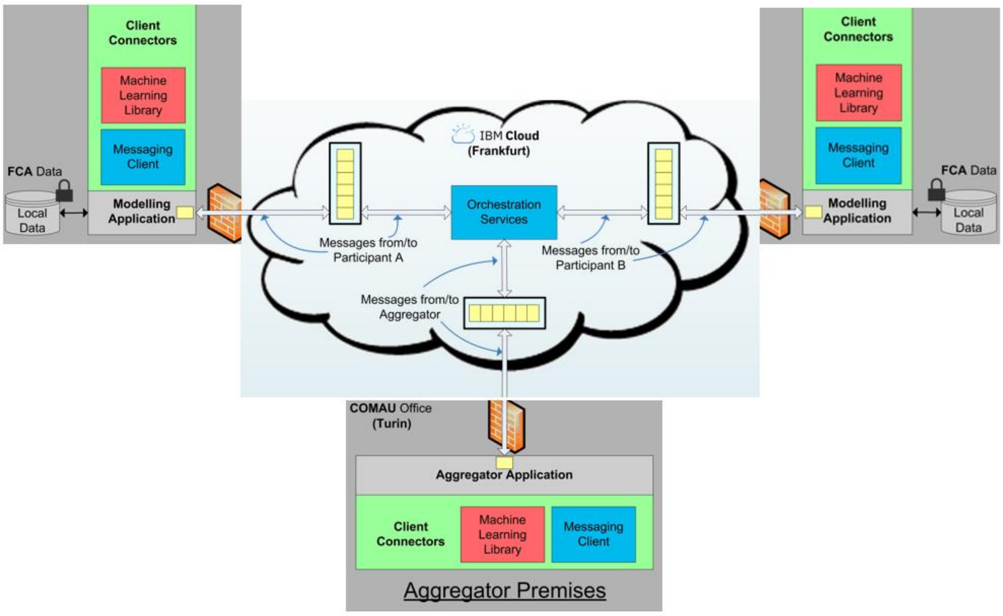
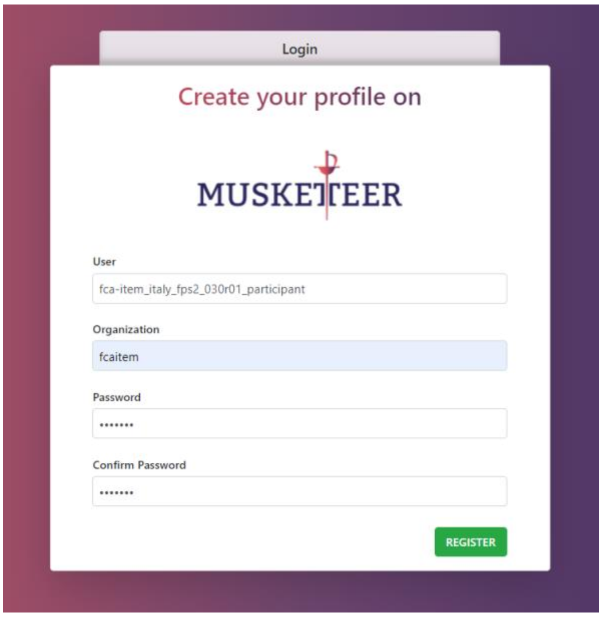


Figure 13 - Pilot platform architecture

The users install the Client Connector using a specific file (called *docker-compose*) provided by ENGINEERING, enabling the microservices interactions.

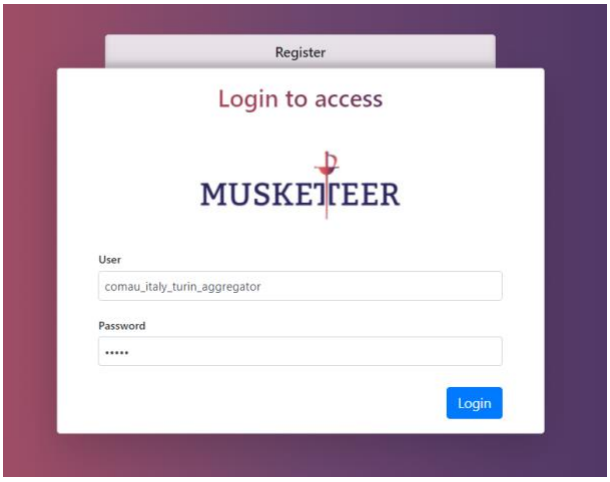
To execute the use case, the first step is to create a user profile on the client connector. Both the aggregator and all the participants do this step with the interface shown below.

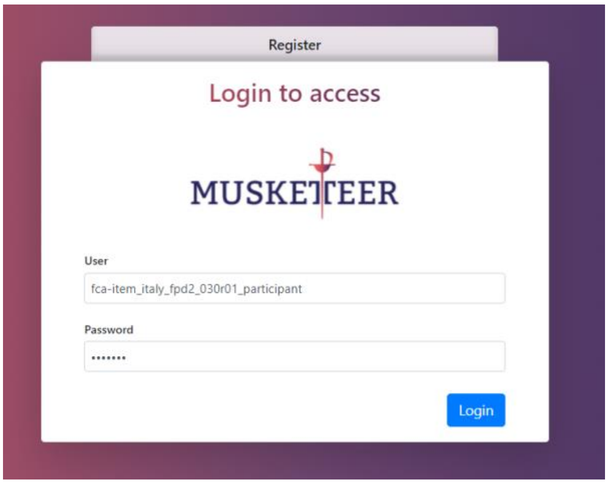


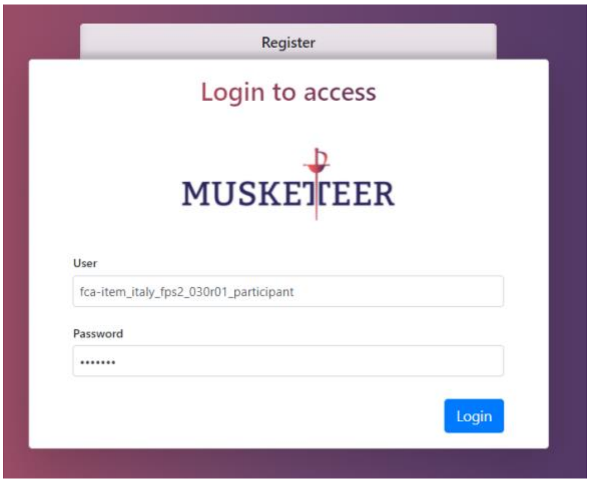
In the Smart Manufacturing use case, COMAU plays the role of the aggregator and the user created is called *comau_italy_turin_aggregator.* FCA-Item, instead, plays both the role of the first and second participant with users *fca-item_italy_fpd2_030r01_participant* and *fca-*

item_italy_fps2_030r01_participant. These two datasets represent data from different plants and there are confidentiality constraints to be enforced between these two users.

In the second step, the aggregator and all the participants, login to the platform on their own local machine. Below the access interface for all of them.

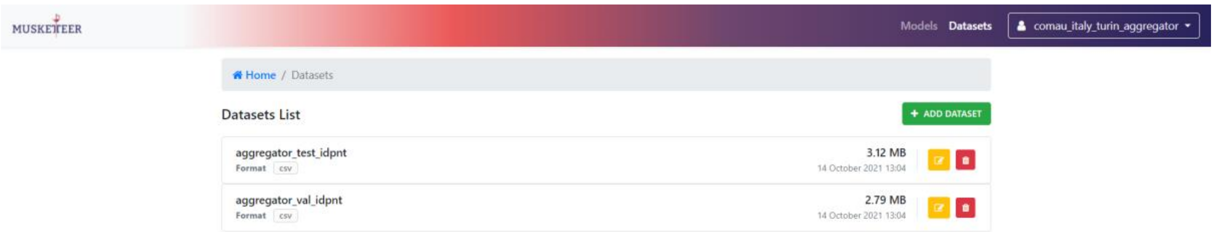




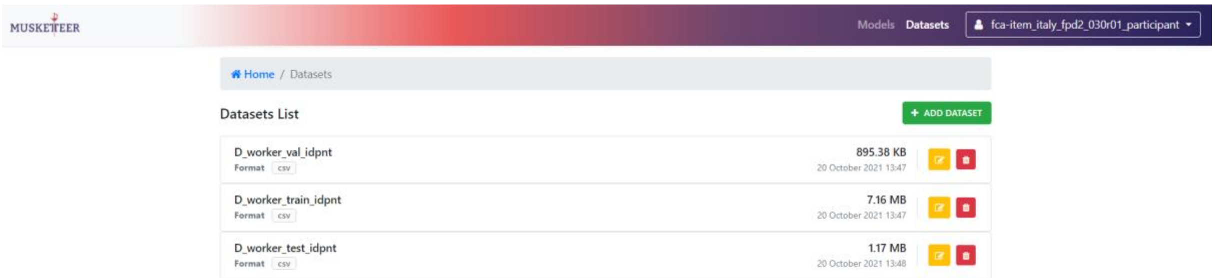


Then, all the players involved locally upload their own datasets. COMAU, as aggregator, has the datasets for test and for validation. The participants have respectively the data relative to the welding of the right side of the vehicle and of the left side. Below the pairs user-dataset available.

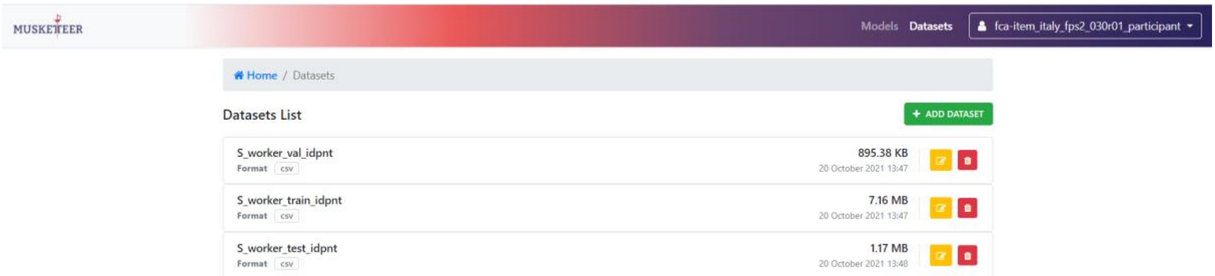
- *comau_italy_turin_aggregator* has the following dataset on his instance of the client connector. These are necessary for the validation and test phase.



- *fca-item_italy_fpd2_030r01_participant* has the following dataset on his instance of the client connector:



- *fca-item_italy_fps2_030r01_participant* has the following dataset on his instance of the client connector:



As previously said, for this use case the chosen algorithm is a Neural Network, to classify the quality of a welding point. To create such a task, the aggregator needs to have a json file that describes the architecture of the net. Essentially, this file characterizes the previously mentioned three layers Neural Network. The entire json used is available on Annex A.

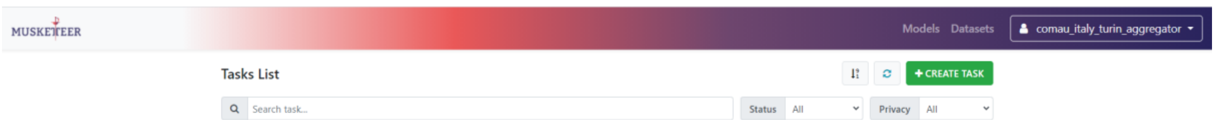
Moreover, to create the classification task, the aggregator needs to have two files to describe the shape of the data: one for the input data and one for the target output. Below, respectively, the two json files.

```
{
      "NI": 5,
      "input_types": [ {
                            "type": "cat",
                "name": "ID_POINT",
                "values":  ["P085",  "P082",  "P086",  "P003",  "P004",
"P001", "P005", "P083", "P081"]
```

```
                }, {
                    "type": "num",
                    "name": "WELCNT"
            }, {
                    "type": "num",
                    "name": "WDRSCNT"
            }, {
                    "type": "num",
                    "name": "WELR"
            }, {
                    "type": "num",
                    "name": "WELCUR1_S"
            }
        ]
}
```

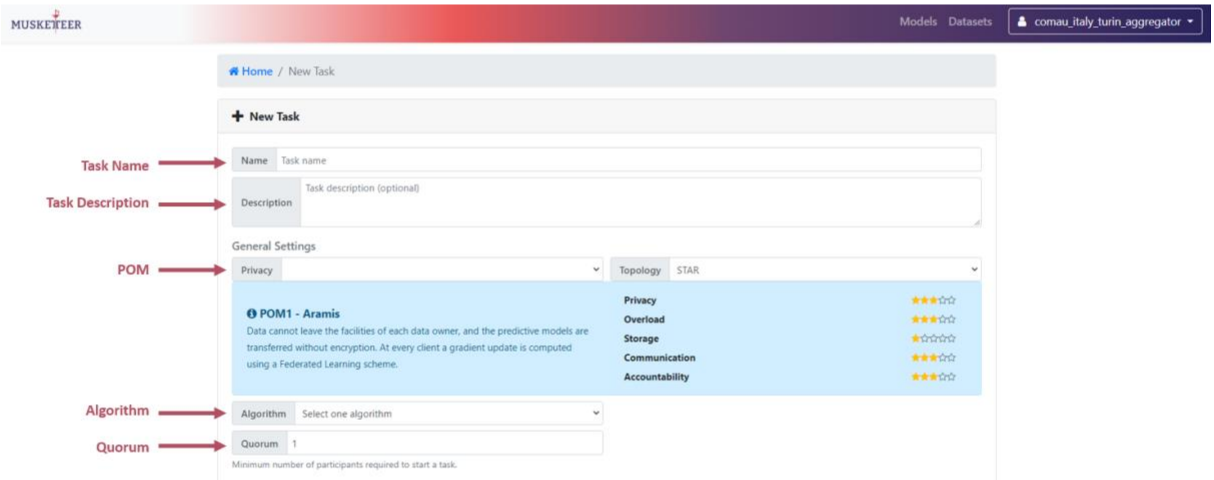
```
{
        "NT": 1,
        "output_types": [{
                    "type": "bin"
            }
        ]
}
```

Once the aggregator inserted all the information needed to create the ML task and these three files, it is ready to create the task. To do so, the user clicks on the button CREATE TASK on the client connector interface, as shown below.

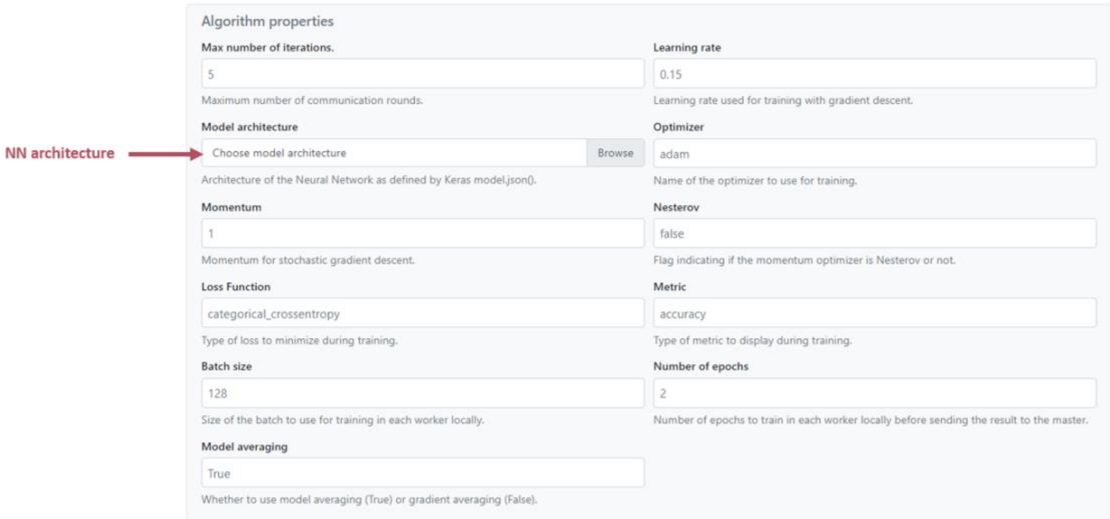


After that, the client connector shows the user a page to characterize the task. The basic information to insert are:

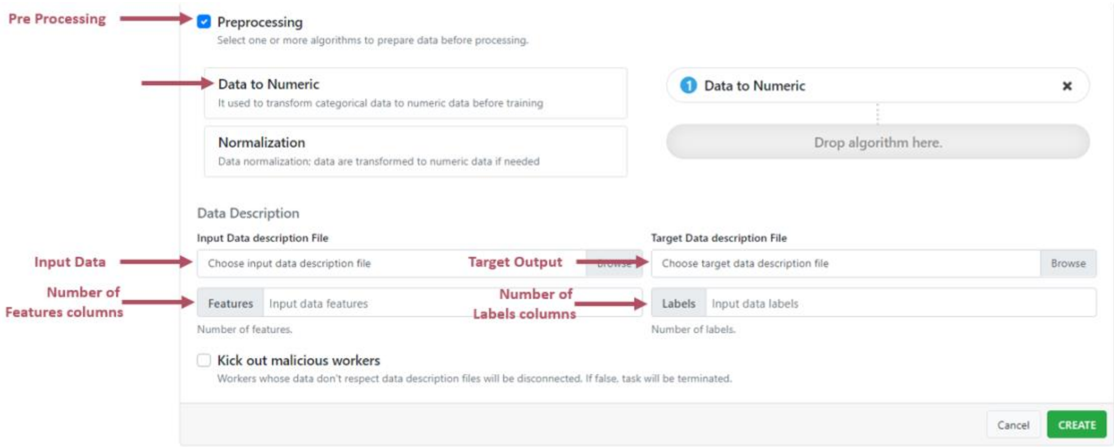
- **Task Name**: the name the user chose for the task.
- **Task Description**: the description the user provides for the task.
- **POM**: the chosen POM.
- **Algorithm**: the chosen algorithm.
- **Quorum**: the expected number of participants.



When the aggregator chooses the POM and the algorithm, the interface proposes the algorithm specific properties needed to proceed. In particular, here the aggregator adds the json file previously described, to characterize the Neural Network chosen.

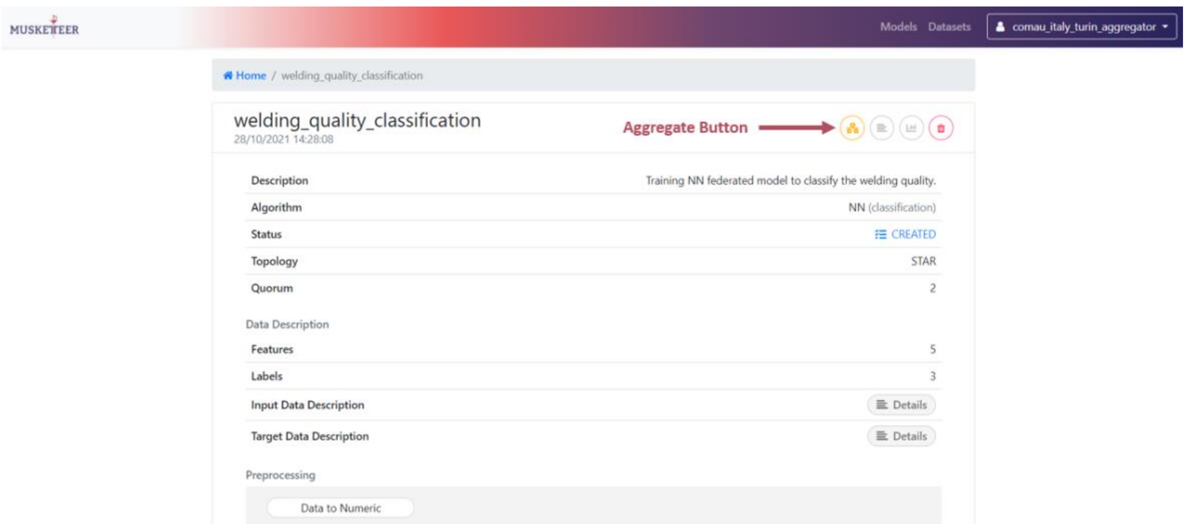


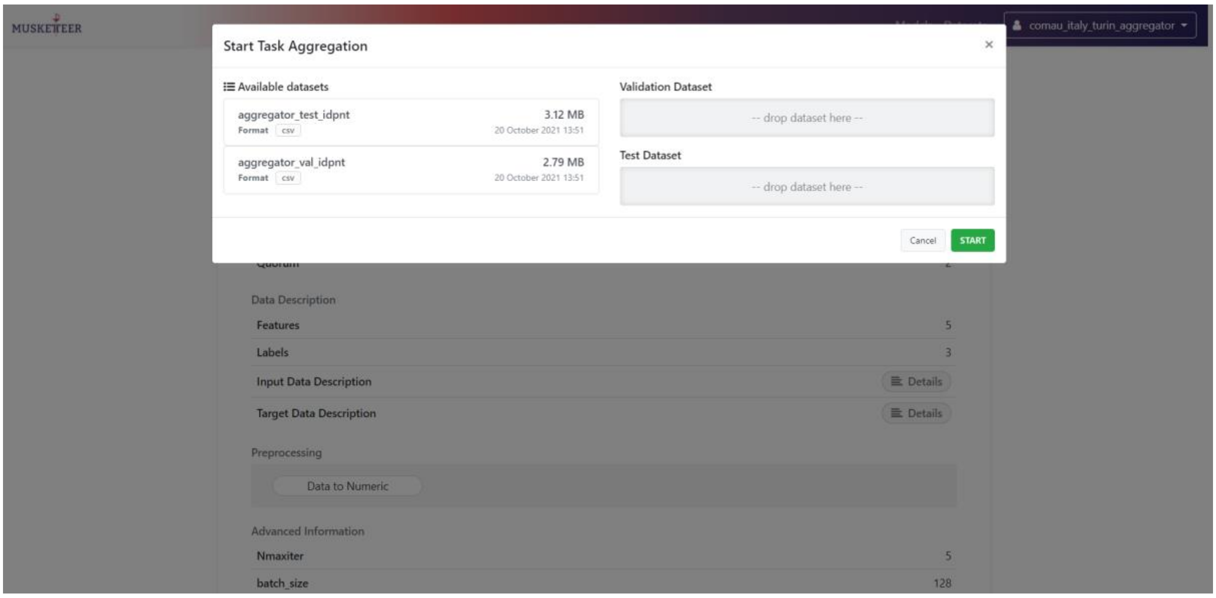
Then, the aggregator chooses to preprocess the data, in particular, in this use case, the data contains a categorical column that has to be transformed into Numeric. Finally, COMAU uploads the description of input and output data and inserts the number of Features columns and the number of Labels columns. Below the interface proposed by the client connector.



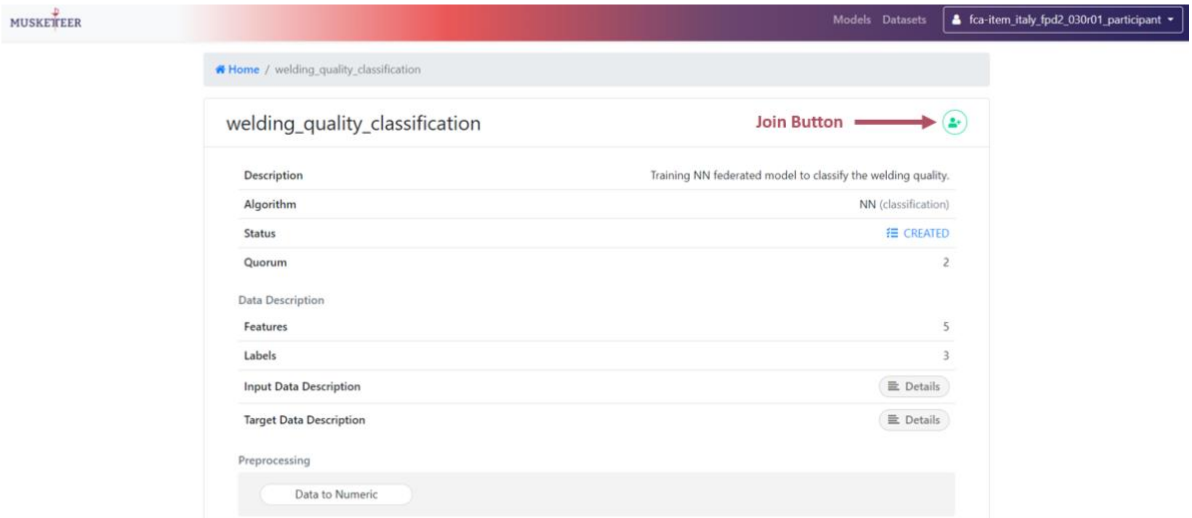
In the next chapter all the detailed characteristics of the tasks done to validate MUSKETEER platform will be specified.

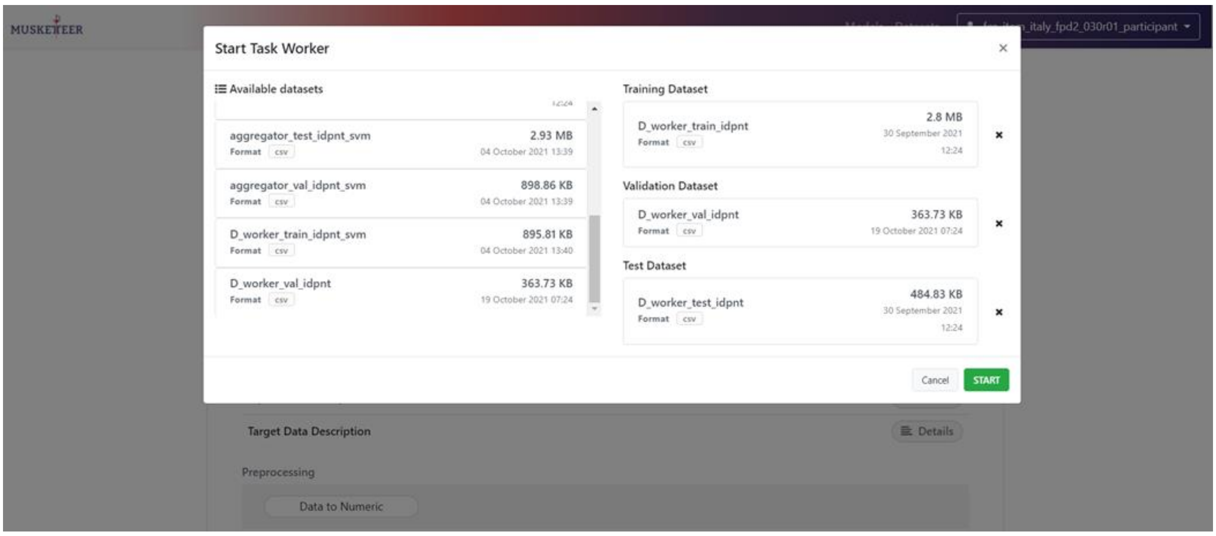
However, when the task is created, COMAU can revise the task and its characteristics. Then it can aggregate it using its two datasets.



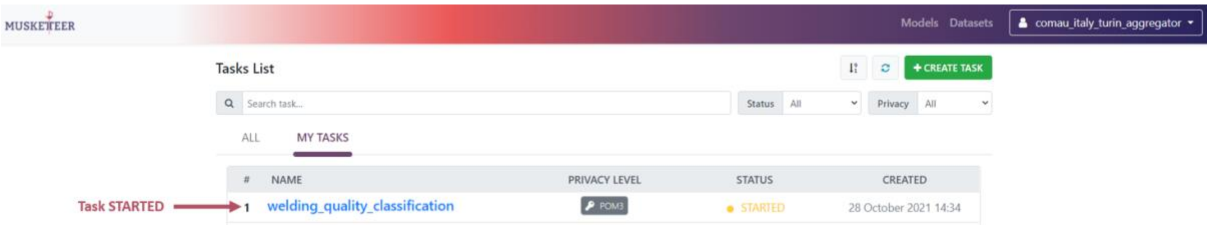


When the task is aggregated, all the participants can join it with their datasets.

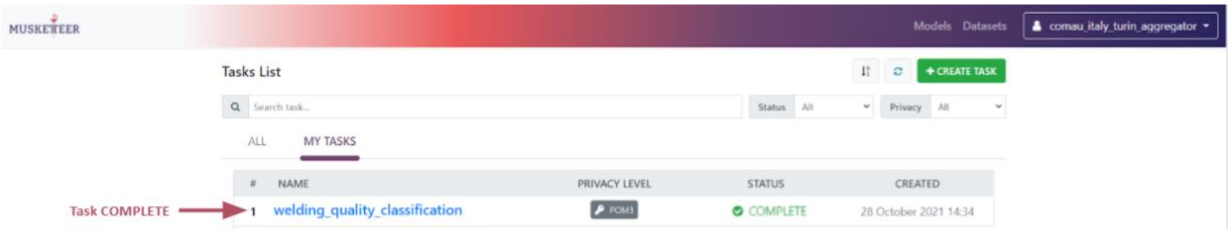




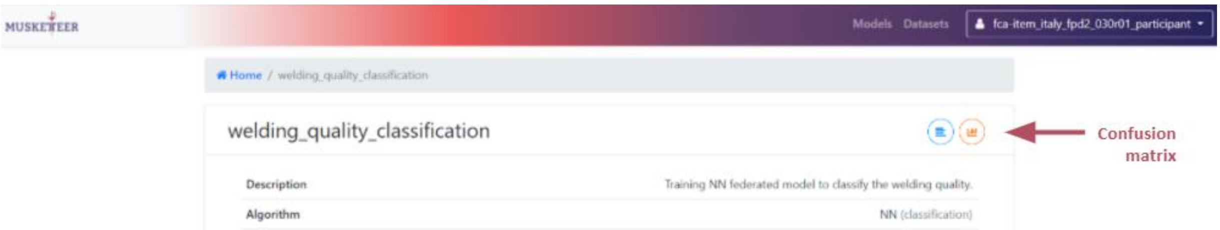
Finally, when the quorum is reached (the expected number of participants has joined the task), the task changes status from CREATED to STARTED.



When all the iterations are done, the task changes for the last time status into COMPLETE.



When the task is in COMPLETE state, on the participant side and id they provided a test dataset, it is available a confusion matrix to evaluate the algorithm performance.



4.6 Results

To validate both the platform and the Federated Machine Learning effectiveness, three tasks are created and executed. The first two have quorum equal to 1 and represent a non-federated approach in which, respectively, the data of a single participant is used to train the algorithm. Instead, the third task, with quorum 2, represents the federated approach in which different participants contribute to increase the overall accuracy of the model. As expected, the results are promising and show its increment.

Below the definition of the first task executed.

Task definition					
Task Name	welding_data_classification_S				
Description	Training NN federated model to classify the welding quality.				
POM	POM3	**Algorithm**	Neural Network	**Quorum**	1
Max Iterations	100	**Pre-processing**	Data to Numeric		

Task execution			
Aggregator	User	comau_italy_turin_aggregator	
	Val dataset	aggregator_val_idpnt.csv	
	Test dataset	aggregator_test_idpnt.csv	
Participants	User	fca-item_italy_fps2_030r01_participant	
		Train dataset	S_worker_train_idpnt.csv
		Val dataset	S_worker_val_idpnt.csv
		Test dataset	S_worker_test_idpnt.csv

The accuracy of the task, made available by the platform on the participant side, is reported in the following table.

Task Results	
Accuracy fca-item_italy_fps2_030r01_participant (Non-federated/Local)	72%

Moreover, to show the task result it is also available on the interface a confusion matrix. This is a specific table layout that allows visualization of the performance of an algorithm. On the horizontal axis there are predicted labels and on the vertical axis the true labels. On each cell *(x,y)* is reported the percentage of instances predicted by the algorithm with label *x* that really have label *y*. The higher are the values on the diagonal, the higher is the quality of the prediction. Notice that, this matrix is calculated using the data available in the test dataset on participant side.

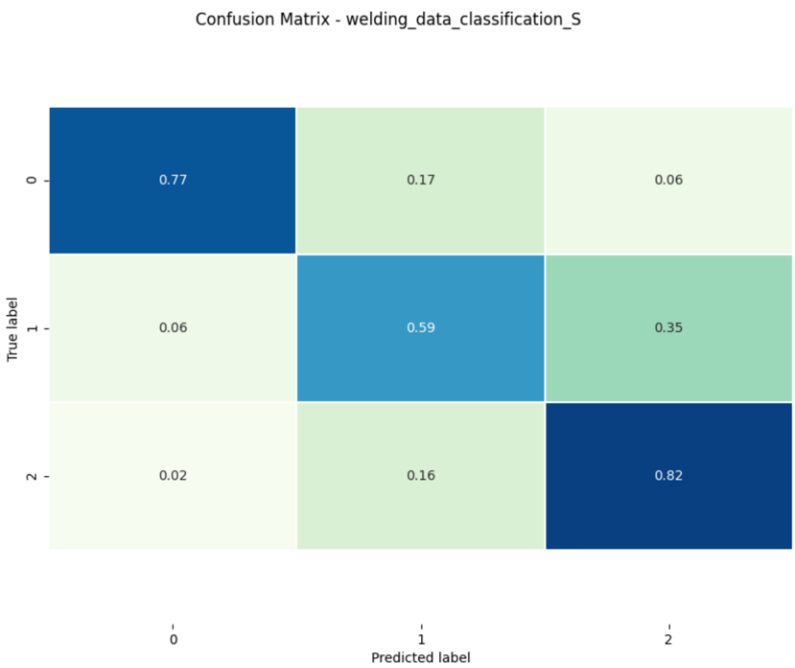


Fig. Confusion Matrix of fca-item_italy_fps2_030r01_participant

Below the definition of the second task executed.

Task definition					
Task Name	welding_data_classification_D				
Description	Training NN federated model to classify the welding quality.				
POM	POM3	**Algorithm**	Neural Network	**Quorum**	1
Max Iterations	100	**Pre-processing**	Data to Numeric		
Task execution					
Aggregator	**User**	comau_italy_turin_aggregator			
	Val dataset	aggregator_val_idpnt.csv			
	Test dataset	aggregator_test_idpnt.csv			
Participants	**User**	fca-item_italy_fpd2_030r01_participant			
		Train dataset	D_worker_train_idpnt.csv		
		Val dataset	D_worker_val_idpnt.csv		
		Test dataset	D_worker_test_idpnt.csv		

Also, for this task the related confusion matrix (on the participant side) is presented below.

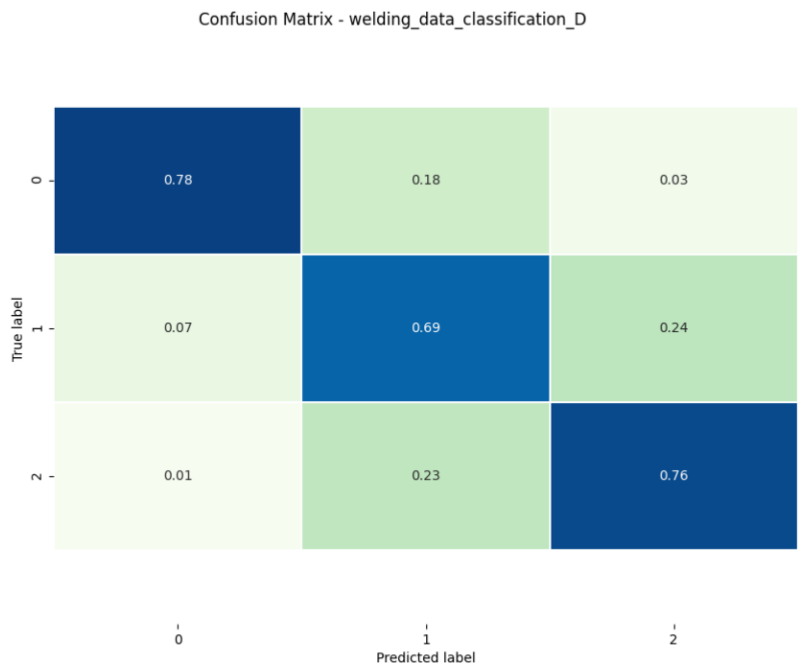


Fig. Confusion Matrix of fca-item_italy_fpd2_030r01_participant

The accuracy of the task available on the participant side is reported in the following table.

Task Results	
Accuracy fca-item_italy_fpd2_030r01_participant (Non-federated/Local)	73%

Finally, the definition of the last task executed with a federated approach.

Task definition					
Task Name	welding_data_classification				
Description	Training NN federated model to classify the welding quality.				
POM	POM3	**Algorithm**	Neural Network	**Quorum**	2
Max Iterations	100	**Pre-processing**	Data to Numeric		
Task execution					
Aggregator	**User**	comau_italy_turin_aggregator			
	Val dataset	aggregator_val_idpnt.csv			
	Test dataset	aggregator_test_idpnt.csv			
Participants	**User**	fca-item_italy_fps2_030r01_participant			
		Train dataset	S_worker_train_idpnt.csv		

		Val dataset	S_worker_val_idpnt.csv
		Test dataset	S_worker_test_idpnt.csv
	User	fca-item_italy_fpd2_030r01_participant	
		Train dataset	D_worker_train_idpnt.csv
		Val dataset	D_worker_val_idpnt.csv
		Test dataset	D_worker_test_idpnt.csv

The result of this task execution shows the increase of accuracy reached with a federated approach with respect to a non-federated one. Below the two confusion matrices of the participants. In both the cases, the one obtained in the federated approach shows this increment.

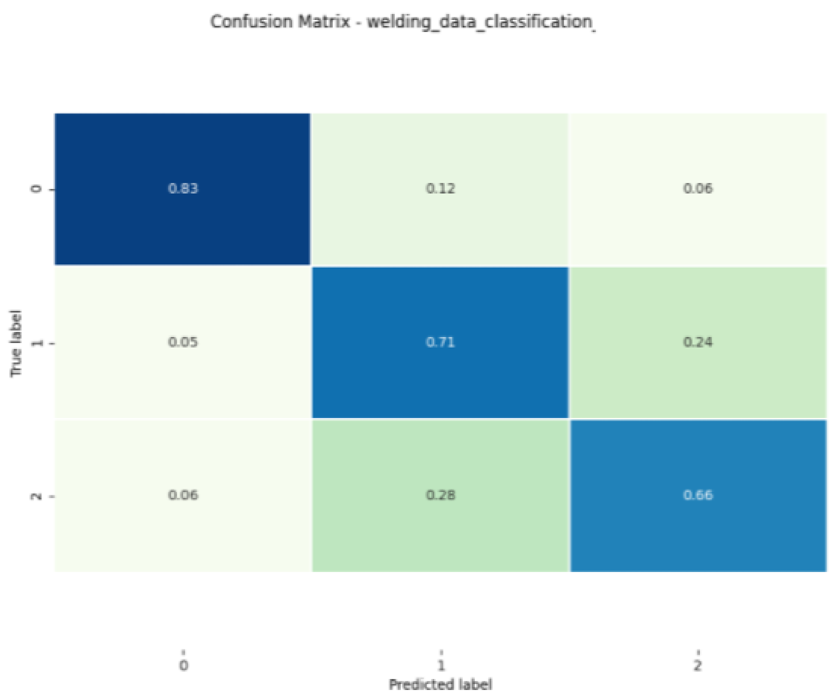


Fig. Confusion Matrix of fca-item_italy_fps2_030r01_participant

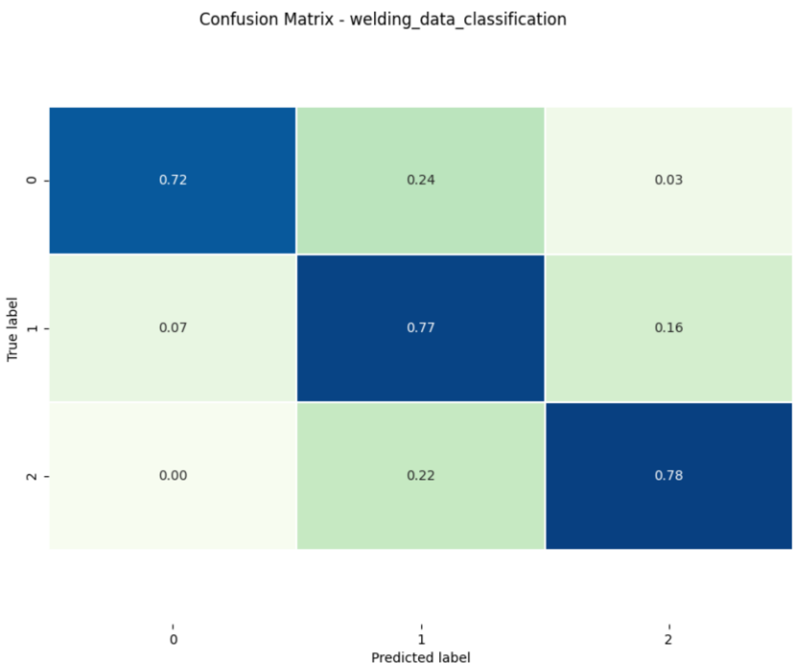


Fig. Confusion Matrix of fca-item_italy_fpd2_030r01_participant

The accuracy of the final task is reported in the following table, compared with the value reached in the non-federated tasks. Both participants obtain a higher accuracy using a federated approach, as expected.

Task Results		
	Non-Federated	Federated
Accuracy fca-item_italy_fps2_030r01_participant	72%	75%
Accuracy fca-item_italy_fpd2_030r01_participant	73%	77%

5 Measurement Goals, Questions and Metrics

This section lists the goals, questions and metrics for the evaluation of the MUSKETEER results that are measured in the Smart Manufacturing use cases.

G1.3	
Analyse	MUSKETEER Architecture
For the purpose of	Evaluate
With respect to	Standardization and extensibility
From the viewpoint of	Business Perspective
In the context of	Smart Manufacturing evaluation scenario (WP7)

The questions identified for the goal G1.3 are listed below.

Identifier	Questions
G1.3_Q01	Does it allow fast deployment, installation and updating?
G1.3_Q02	Is it easy to use?
G1.3_Q03	Are there different visibility constraints based on user permissions?
G1.3_Q04	Is the architecture compliant with industry standard and production plant IT policies?
G1.3_Q05	Does the platform require a special hardware locally?
G1.3_Q06	Is it possible to download the model?
G1.3_Q07	Is the training of the model fast enough?
G1.3_Q08	When a new task is launched, what is the algorithm used and its parameters?
G1.3_Q09	Is it possible to report a comment on an unexpected behavior of algorithm during a *business* user session?

G2.3	
Analyse	Privacy Preserving Operation Modes
For the purpose of	Evaluate
With respect to	Privacy, computational overload, central storage requirements, communication requirements, data utility accountability
From the view point of	Business Perspective
In the context of	Smart Manufacturing evaluation scenario (WP7)

The questions identified for the goal G2.3 are listed below.

Identifier	Questions
G2.3_Q01	How easy it is to declare the privacy requirements?

G2.3_Q02	Due to our policy, is an adequate level of data privacy granted?
G2.3_Q03	All the features of the selected POM are implemented?
G2.3_Q04	How easy it is to encrypt or decrypt data or model?
G2.3_Q05	Does my central storage support the platform requirements?
G2.3_Q06	How easy it is to verify if all the communications are working?
G2.3_Q07	Which is the maximum dimension of messages supported by the platform?

G3.4	
Analyse	Machine Learning Algorithms
For the purpose of	Evaluate
With respect to	pre-processing, normalization, data alignment, supervised and unsupervised learning
From the view point of	Business Perspective
In the context of	Smart Manufacturing evaluation scenario (WP7)

The questions identified for the goal G3.4 are listed below.

Identifier	Questions
G3.4_Q01	Given historical data of the welding gun (current, welding time, welding force….) and a model, is MUSKETEER able to provide the improvement for grouping the dataset in such a way that objects in the same group are more similar to each other than to those in other groups?
G3.4_Q02	Given the data on the welding gun (current, welding time, welding force….) and a model, is MUSKETEER able to improve the prediction of future *n* values of quality index?
G3.4_Q03	Given the data on the welding gun (current, welding time, welding force….) and a model, is MUSKETEER able to improve the classification of quality index class?
G3.4_Q04	Are the results of the prediction interpretable? In particular, when the task is completed, is the outcome characterized also with all information concerning the context (a header containing model, algorithm, …) where it has been executed?
G3.4_Q05	Is the algorithm able to use data of different plants to extract knowledge useful for all?
G3.4_Q06	Is the ML algorithm reliable? Does it give comparable output working on the same data and in the same conditions in different sessions?
G3.4_Q07	Does the prediction provided by the model (after the MUSKETEER training) provide an improved knowledge to build a reliable relationship between spot weld issues and possible causes linked to applied welding process (parameters values, …)?

5.1 Metrics

All the metrics belonging to the smart manufacturing use case are listed below.

These metrics allow to estimate the ease of installation and use of the platform, the time and the cost spent and the evaluation of the obtained result with Smart Manufacturing pilot.

The provided evaluation is the average of the KPIs estimated by FCA and the ones estimated by COMAU (SMEG).

Identifier	KPI	Format	Method of collection and measurement	Evaluation
G1.3_Q01_M01	(Time taken to deploy and install the MUSKETEER client) * (Number of employees involved to deploy and install the MUSKETEER client)	(HH:MM) * number of employees	Online questionnaire; face-to-face interview	02:00*1
G1.3_Q01_M02	(Time taken to update the MUSKETEER client) * (Number of employees involved to update the MUSKETEER client)	(HH:MM) * number of employees	Online questionnaire; face-to-face interview	00:30*1
G1.3_Q02_M01	Time taken by one person to create a task	HH:MM	Online questionnaire; face-to-face interview	00:02
G1.3_Q02_M02	Time taken to run the training procedure associated to a given ML task / training data size	HH:MM	Online questionnaire; face-to-face interview	From 00:05 to 01:00 depending by the POM
G1.3_Q02_M03	Time taken by one person to add a new user	HH:MM	Online questionnaire; face-to-face interview	00:02
G1.3_Q03_M01	Different visualizations for different user permissions	Boolean (true/false)	Verification of platform	True
G1.3_Q04_M01	Compliance with Information Technology production plant policies (used ports, protocols constraints, ...)	Boolean (true/false)	Verification of platform technical documentation	False

G1.3_Q05_M01	Cost of local special equipment	Number (EUR)	Online questionnaire; face-to-face interview	0€
G1.3_Q05_M02	Cost of setting up local special equipment	Number (EUR)	Online questionnaire; face-to-face interview	0€
G1.3_Q06_M01	Possibility of downloading the ML model	Boolean (true/false)	Availability of a button to download the ML model	True
G1.3_Q07_M01	Time taken to classify the current status of the equipment	HH:MM	Verification of platform	00:01
G1.3_Q07_M02	Time taken to classify the future status of the equipment	HH:MM	Verification of platform	00:01
G1.3_Q08_M01	Possibility to see a resume page to show what is the algorithm used and its parameters	Boolean (true/false)	Verification of platform	True
G1.3_Q09_M01	Report an unexpected behavior	Boolean (true/false)	Availability of a comment area linked to the session *Measurement*: Availability of area	True
G2.3_Q01_M01	Time taken by one person to declare privacy requirements	HH:MM	Online questionnaire; face-to-face interview	00:01
G2.3_Q01_M02	Presence of a clear description of each POM and its implications	Boolean (true/false)	Verification of platform	True
G2.3_Q01_M03	Number of options expressed in natural language / total number of steps	Decimal	Online questionnaire; face-to-face interview	1
G2.3_Q02_M01	Possibility of sharing my data with other users	Boolean (true/false)	Online questionnaire; face-to-face interview	False
G2.3_Q02_M02	Possibility to control if other users are visualizing my data	Boolean (true/false)	Online questionnaire; face-to-face interview	False

G2.3_Q02_M03	Possibility to control who has the grant for visualizing my data	Boolean (true/false)	Online questionnaire; face-to-face interview	False
G2.3_Q03_M01	Correct receipt of my private key	Boolean (true/false)	Online questionnaire; face-to-face interview	True (log file)
G2.3_Q03_M02	Correct encryption of data or model using my private key	Boolean (true/false)	Real time test between end users and central node to compare data or model	True (log file)
G2.3_Q03_M03	Correct receipt of central node encrypted results of computation	Boolean (true/false)	Online questionnaire; face-to-face interview	True (log file)
G2.3_Q03_M04	Correct decrypt of central node results	Boolean (true/false)	Real time test between end users and central node to compare data or model	True (log file)
G2.3_Q04_M01	Time taken by the POM selected to encrypt or decrypt data or model	HH:MM	Online questionnaire; face-to-face interview	From 00:02 to 00:15 depending by the POM
G2.3_Q05_M01	(my central -*local for end user*- storage size) - (Storage requested by the platform)	Decimal	Direct calculation performed by the platform. Measurements: KPI has to be > 0	7.5GB
G2.3_Q06_M01	Possibility to verify if all the communication protocols are enabled	Boolean (true/false)	Online questionnaire; face-to-face interview	True
G2.3_Q07_M01	Maximum dimension of data loaded in the platform	Decimal	Stress tests	18.7 MB for Smart Manufacturing use case
G3.4_Q01_M01	Availability of cluster cohesion index when a clustering model has been chosen	Boolean (true/false)	Clustering analysis with *n* groups using all main welding parameters	True: available a similar KPI: “Average distance to the closest centroids”.
G3.4_Q02_M01	Mean squared error of the difference between	Double	Test ML models	0.1

	the predicted quality index and the real quality index			
G3.4_Q03_M01	Accuracy of the classification when a supervised model has been chosen	Percentage	Test ML models	The accuracy value is available; ~80% in the pilot case
G3.4_Q03_M02	Precision and recall measure when a supervised model is chosen: Precision = TRUE POSITIVE / (TRUE POSITIVE + FALSE POSITIVE) Recall = TRUE POSITIVE / (TRUE POSITIVE + FALSE NEGATIVE)	Double	Test ML models	Values calculated in the pilot case: Precision: 76% Recall: 77%
G3.4_Q04_M01	Time taken to one person to understand algorithm output	HH:MM	Online questionnaire; face-to-face interview	00:05
G3.4_Q04_M02	Return of the parameters which most influence the model	Boolean (true/false)	Online questionnaire; face-to-face interview	False
G3.4_Q04_M03	Completeness of output header	Boolean (true/false)	Check list of fields to be characterized *Measurement*: 100% of fields have a value	True
G3.4_Q05_M01	Accuracy of federated model >= accuracy of local (trained with single plant's data) model	Boolean (true/false)	Test ML model	True
G3.4_Q06_M01	Difference of two output calculated on same input in different sessions	Double	Comparison of output *Measurement:* zero or not relevant differences detected	Not relevant differences detected
G3.4_Q06_M02	Ratio of data errors in pre-processing,	Percentage	Tests in pre-processing,	0%

	normalization and data alignment		normalization and data alignment	
G3.4_Q06_M01	Provide a cause-effect relationship	Two-dimensional array whose values are boolean	Face to face interview *Measurements*: Presence of “true” values in two-dimensional array	Not Available

Table 2 -Smart manufacturing metrics

6 The Technical Perspective

G1.1	
Analyse	MUSKETEER Architecture
For the purpose of	Evaluate
With respect to	Standardization and extensibility
From the view point of	Technical Perspective
In the context of	Use Cases validation (WP7)

The questions identified for the goal G1.1 are listed below.

Identifier	**Questions**
G1.1_Q01	Is the MUSKETEER architecture aligned with the Industrial Data Space Association reference architecture?
G1.1_Q02	Does it allow interoperability with ML frameworks?
G1.1_Q03	Does it foster the creation of a community of developers and researchers that can extend the platform with new algorithms and attack detection mechanisms?
G1.1_Q04	Does it allow fast deployment, installation and use?

G2.1	
Analyse	Privacy Preserving Operation Modes
For the purpose of	Evaluate
With respect to	Privacy, computational overload, central storage requirements, communication requirements, data utility accountability
From the view point of	Technical Perspective
In the context of	Use Cases execution (WP7)

The questions identified for the goal G2.1 are listed below.

Identifier	Questions
G2.1_Q01	Will POMs be designed to allow a secure information exchange among platform user?
G2.1_Q02	Will POMs provide compliance with the legal and confidentiality restrictions of most industrial scenarios?
G2.1_Q03	Will the scalability of ML algorithms be improved over every POM, with regard to correct combination of different concepts of federated ML, differential privacy, homomorphic encryption, secure multiparty computation and distributed computing?

6.1 Technical perspective metrics

Identifier	KPI	Format	Method of collection and measurement	Evaluation
G1.1_Q01_M01	Number of artifacts compliant with IDSA reference architecture / Total number of artifacts	Decimal <=1	Online questionnaire; face-to-face interview	0.86
G1.1_Q02_M01	Number of ML libraries supported to export the predictive models /Total of the best-known ML libraries	Decimal <=1	Online questionnaire; fate-to-face interview	1
G1.1_Q03_M01	Open-source web communities' interactions	Integer	Field survey	4
G1.1_Q0.4_M01	Number of SW applications released as images	Integer	Field survey	4
G1.1_Q0.4_M02	Number of software components released in open-source repositories	Integer	Field survey	6
G2.1_Q01_M01	Number of robust POMs for use cases	Integer	On field	6
G2.1_Q02_M01	Speedup/number of users while POM is applied	Ratio	On field	POM dependent, see D6.2

G2.1_Q03_M01	Number of training procedures implemented	Integer	On field	33
G2.1_Q03_M02	Speed of privacy-preserving machine learning algorithms implemented with respect to other existing solutions	Ratio	On field	Not computable, see D6.2

Table 3 – Technical perspective metrics

7 Conclusion

This deliverable shows the results of the instantiation of the Smart Manufacturing use case in MUSKETEER platform and evaluates the platform respect to business and technical KPIs.

After the recap of the general performance indicators, FCA and COMAU evaluated them as end users of the platform for the Smart Manufacturing use case with the Goal-Question-Metric approach. It emerges that the platform is complete, it offers six different privacy operation modes, different families of ML models and, not less important, it respects the security requirements that were expected.

Then the use case has been detailed and it has been described how the platform can solve the main issues of the pilot.

FCA and COMAU chose POM3-Porthos as privacy operation mode, because the final model is private and not shared between the different participants. The model used in the pilot is a supervised model able to classify the quality of the welding points, using welding parameters as input. A Neural Network has been chosen.

The end users approach the platform thanks to a client connector instance locally installed. Some screenshots have been inserted to guide the reader though the entire procedure of creating a task, train the model and receiving the results and to show the simple and user-friendly interface.

The result of the execution clearly shows that the federated scenario has a better accuracy than a model trained on a single dataset, so the main goal of MUSKETEER project has been reached in this use case.

It is possible to state, in the light of the results, that the use of the solution represents a concrete opportunity to achieve a more efficient maintenance able to intercept, with predictive models, possible future failures or decrease in the quality of the welds in charge of the welding guns.

The reliability and security of the platform and the availability of an adequate privacy operation mode also made it possible to experiment with an operational scenario consistent with the roles in the domain of the FCA and COMAU stakeholders. The pilot highlighted the benefits that can derive from the adoption of federated learning in the industrial context and concerning the improvement of the welding process and the final product associated with it and a reduction in the maintenance costs of the welding guns.

8 References

[1] R. van Solingen, E. Berghout (1999). The Goal/Question/Metric Method: A practical Guide for quality Improvement of Software Development

[2] MUSKETEER Technical Annex

[3] ISO/IEC 25010:2011, System and software engineering – Systems and software quality Requirements and Evaluation (SQuaRE) - System and software quality models

[4] Q.Song, W. Zang, N. Bay (2005) An experimental study determines the electrical contact resistance in resistance welding – Welding journal – vol. 84

9 Annex A

Below the json file to describe the neural network used in the smart manufacturing tasks.


```
{
  "class_name": "Sequential",
  "config": {
    "name": "sequential_1",
    "layers": [
      {
        "class_name": "Dense",
        "config": {
          "name": "dense_1",
          "trainable": true,
          "batch_input_shape": [
            null,
            13
          ],
          "dtype": "float32",
          "units": 48,
          "activation": "relu",
          "use_bias": true,
          "kernel_initializer": {
            "class_name": "VarianceScaling",
            "config": {
              "scale": 1,
              "mode": "fan_avg",
              "distribution": "uniform",
              "seed": null
            }
          },
          "bias_initializer": {
            "class_name": "Zeros",
            "config": {}
          },
          "kernel_regularizer": null,
          "bias_regularizer": null,
          "activity_regularizer": null,
          "kernel_constraint": null,
          "bias_constraint": null
        }
      },
      {
        "class_name": "Dense",
        "config": {
          "name": "dense_2",
          "trainable": true,
          "batch_input_shape": [
            null,
```

```
            5
          ],
          "dtype": "float32",
          "units": 16,
          "activation": "relu",
          "use_bias": true,
          "kernel_initializer": {
            "class_name": "VarianceScaling",
            "config": {
              "scale": 1,
              "mode": "fan_avg",
              "distribution": "uniform",
              "seed": null
            }
          },
          "bias_initializer": {
            "class_name": "Zeros",
            "config": {}
          },
          "kernel_regularizer": null,
          "bias_regularizer": null,
          "activity_regularizer": null,
          "kernel_constraint": null,
          "bias_constraint": null
        }
      },
      {
        "class_name": "Dense",
        "config": {
          "name": "dense_3",
          "trainable": true,
          "units": 3,
          "activation": "softmax",
          "use_bias": true,
          "kernel_initializer": {
            "class_name": "VarianceScaling",
            "config": {
              "scale": 1,
              "mode": "fan_avg",
              "distribution": "uniform",
              "seed": null
            }
          },
          "bias_initializer": {
            "class_name": "Zeros",
            "config": {}
          },
          "kernel_regularizer": null,
          "bias_regularizer": null,
          "activity_regularizer": null,
```

```
          "kernel_constraint": null,
          "bias_constraint": null
        }
      }
    ]
  },
  "keras_version": "2.2.4",
  "backend": "tensorflow"
}
```