Application Of AI (Knowledge Graph Embeddings) Industrial Use Cases Formal Knowledge Integration in Machine Learning Model For Industry 4.0

Industrial PhD Student, **Mouloud IFERROUDJENE**

Supervised by Antoine ZIMMERMANN, Victor CHARPENAY, Thierry LAVEILLE

LABORATOIRE D'INFORMATIQUE, DE MODÉLISATION ET D'OPTIMISATION DES SYSTÈMES

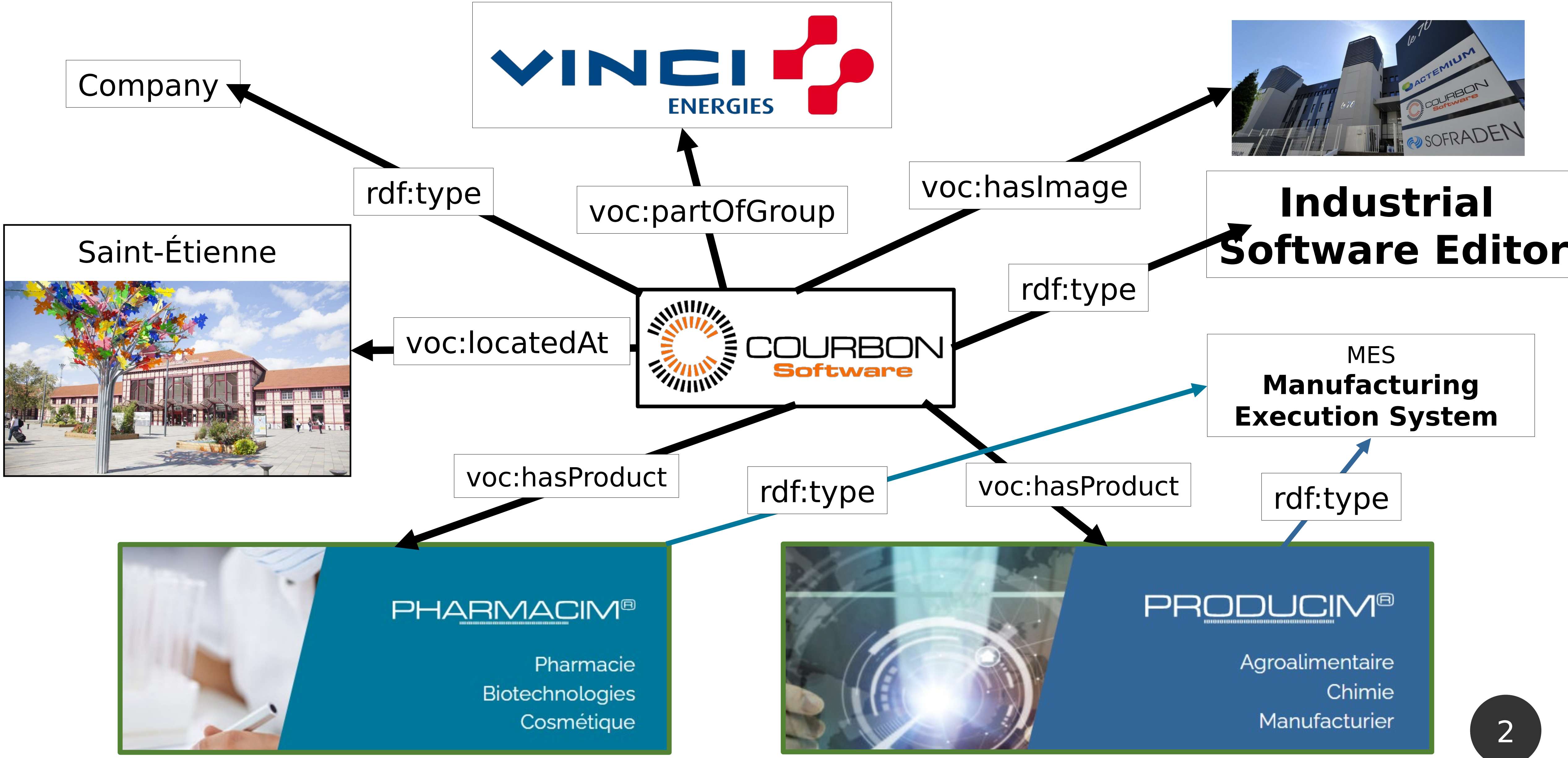
MINES Saint-Étienne

ASSOCIATION NATIONALE RECHERCHE TECHNOLOGIE

COURBON Software

\°/ All Icons used are courtesy of flaticon.com

COURBON Software (CSO)



What is a MES ?

MES Stands for

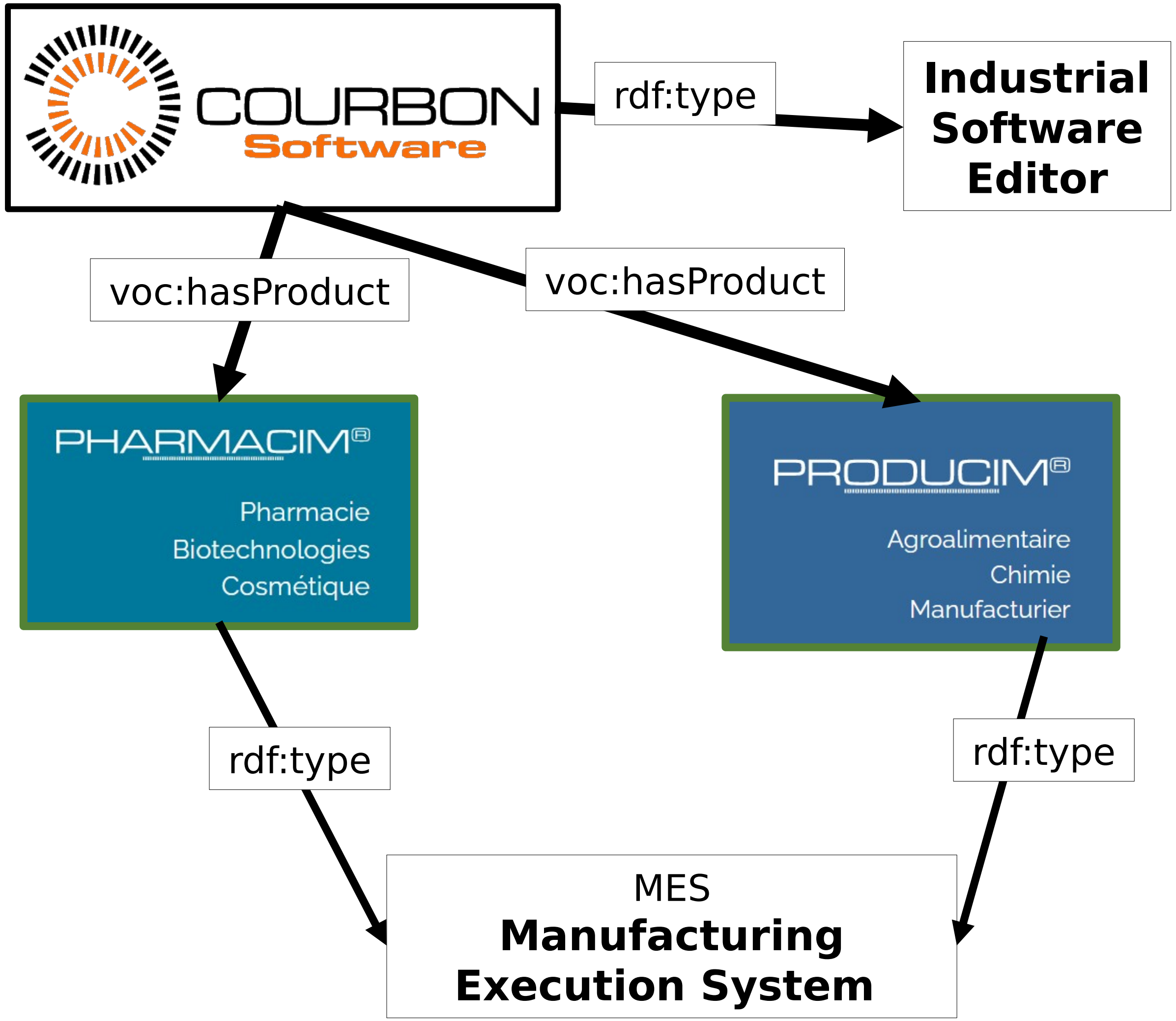
- **Manufacturing execution system**

MES objectives :

- Ensure the proper execution of manufacturing operations
- Improve production efficiency

MES Functionalities :

- Product traceability
- Quality control
- Production monitoring
- Scheduling, Etc.



Data acquisition => Lot of DATA from different sources

CONTEXT

Illustration Example (toy car production)

Imagine you have a toy factory where you make toy cars !!

the past, you had to do everything by hand:

Make the cars

Paint them

And package them

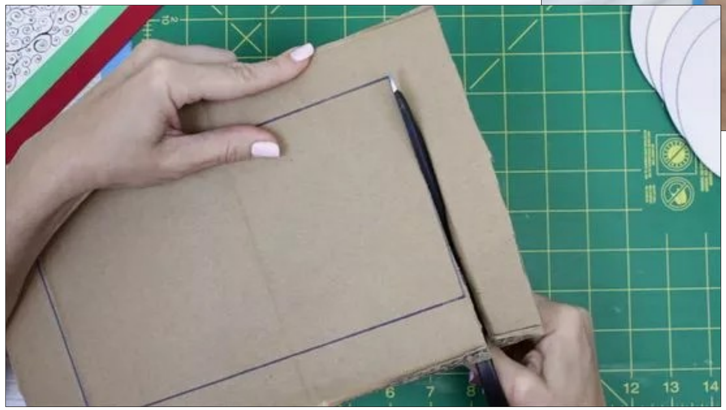
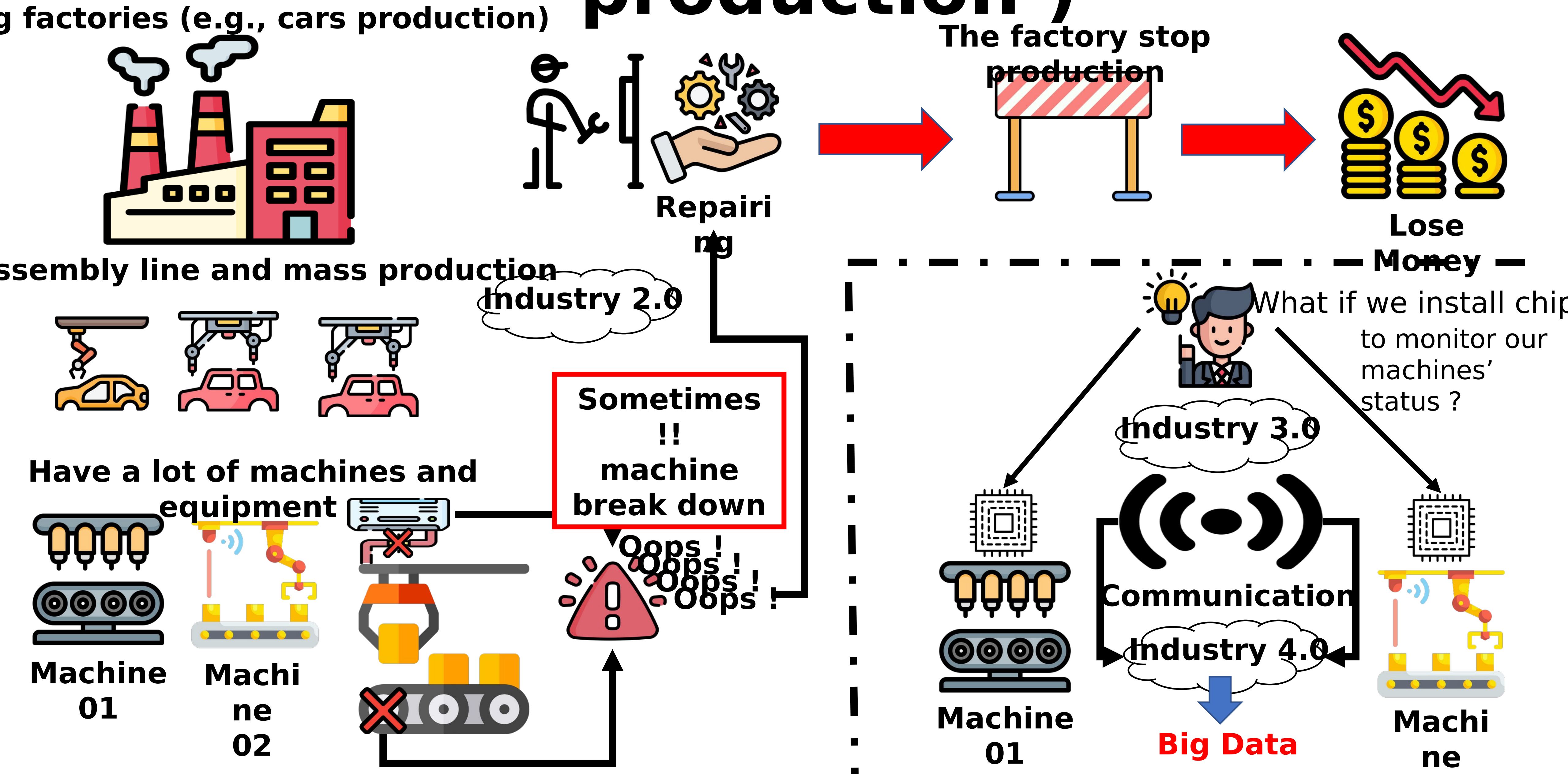
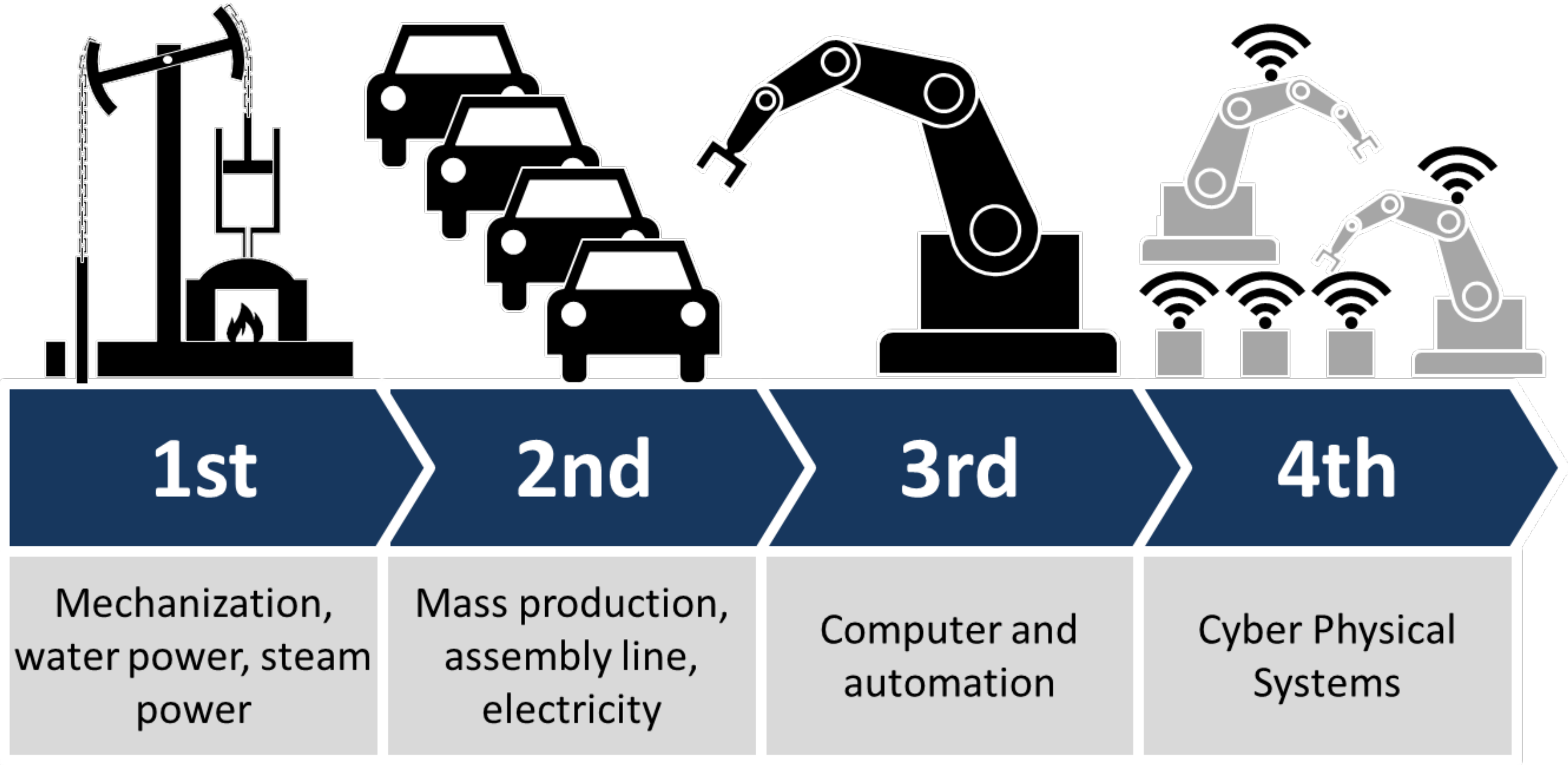


Illustration Example (car toy production)



The Fourth Industrial Revolution (Industry 4.0)

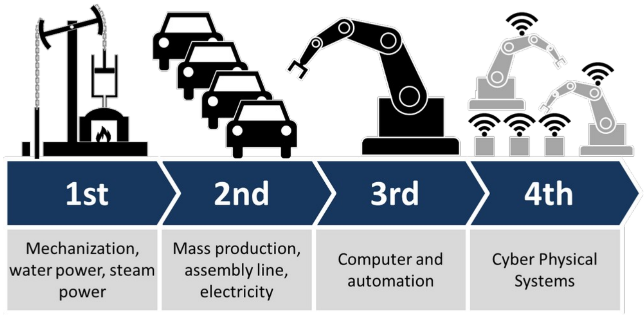


CONTEXT

COURBON Software

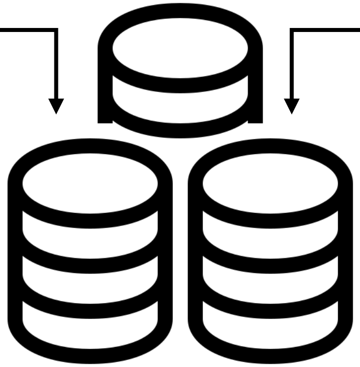
Manufacturing Execution Systems (MES)

Enterprise Resource Planning (ERP)



Industry 4.0

+



Big data

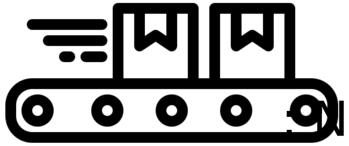
AI applications in the industry :

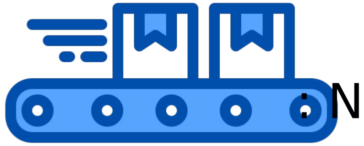
- early detection of rejects
- **Predictive Maintenance**
- quality control
- industrial prognosis
- etc.

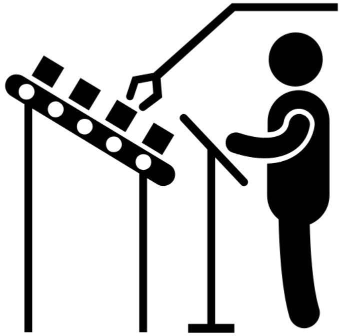
Intelligent industrial system

Labeling of the cause of the error

- Workers labels by hand the cause of failures of every equipment, e.g.,


: Noticed **High Speed** of **the conveyor**


: Noticed **Low energy** of the conveyor



- Every data about the status of machine is collected
- Every issues noticed are reported and data are labled

⇒Extract insight from data to **solve** the **prevent future** the problems.

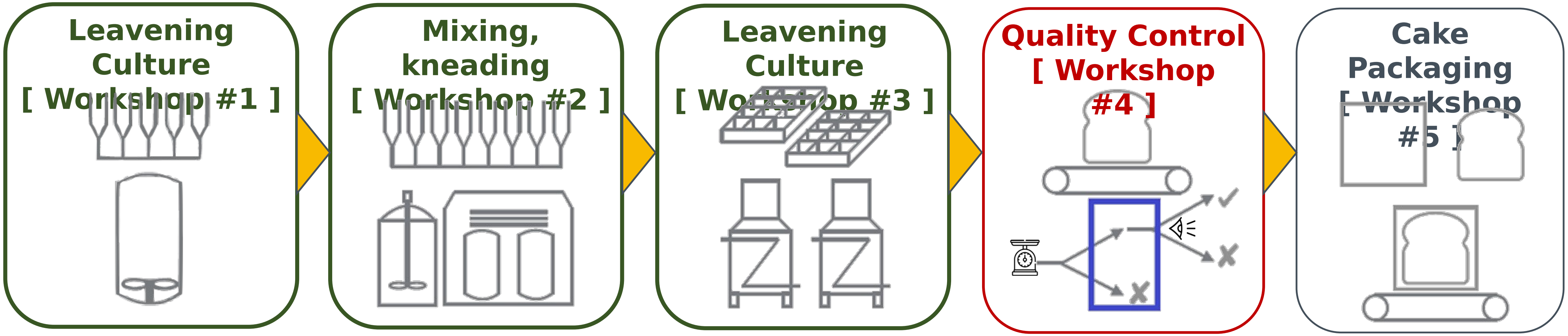
⇒The manager (or workers) gain domain-specific knowledge and expertise.

Create Database

Machine N°	DateTime	Product lot N°	Issue label
M1	24022023T19:00:02	152 (Material)	0
M2 (Conveyor)	24022023T16:02	12 (Toy)	Low Energy

Use Case – Quality Control in Cake Factory

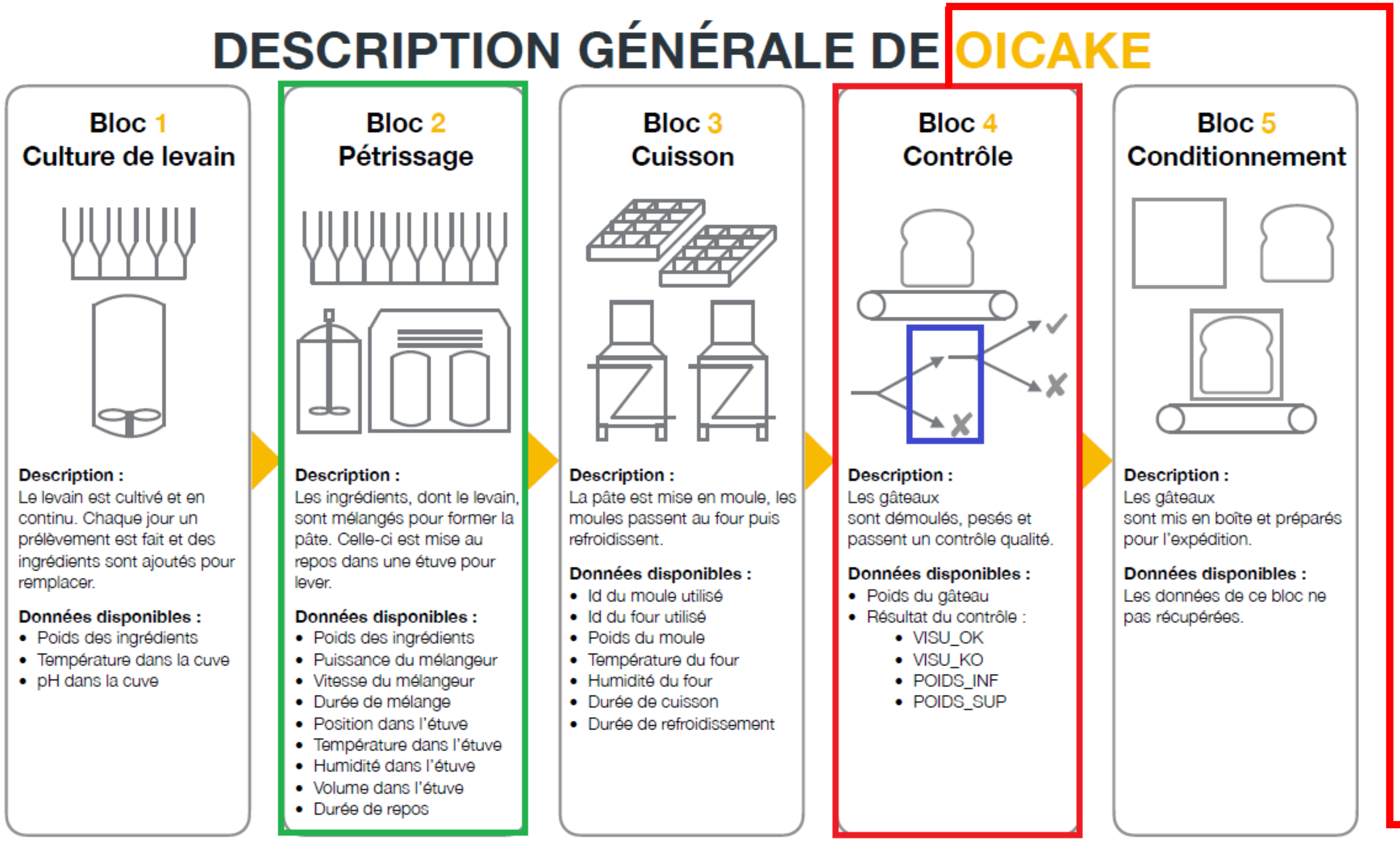
Example of Industrial Use Case



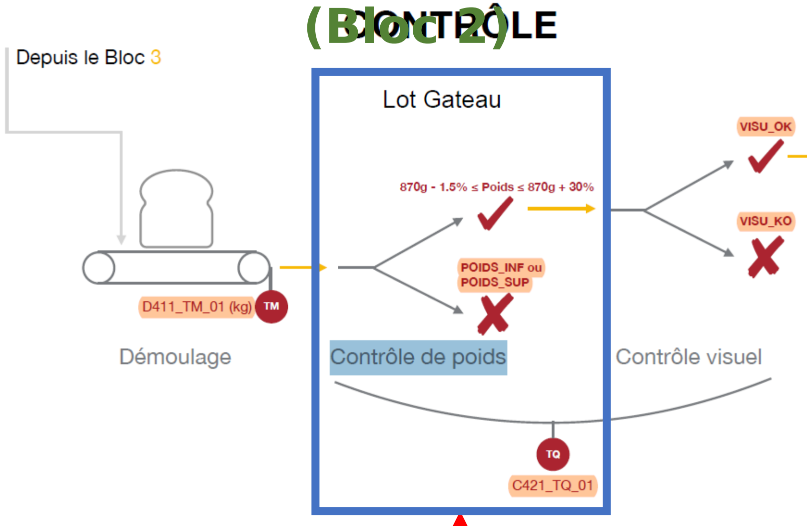
Description of an example **of Sequential Production-Line**

Use Case – Quality Control in Cake Factory

A reminder of OICake production line schema



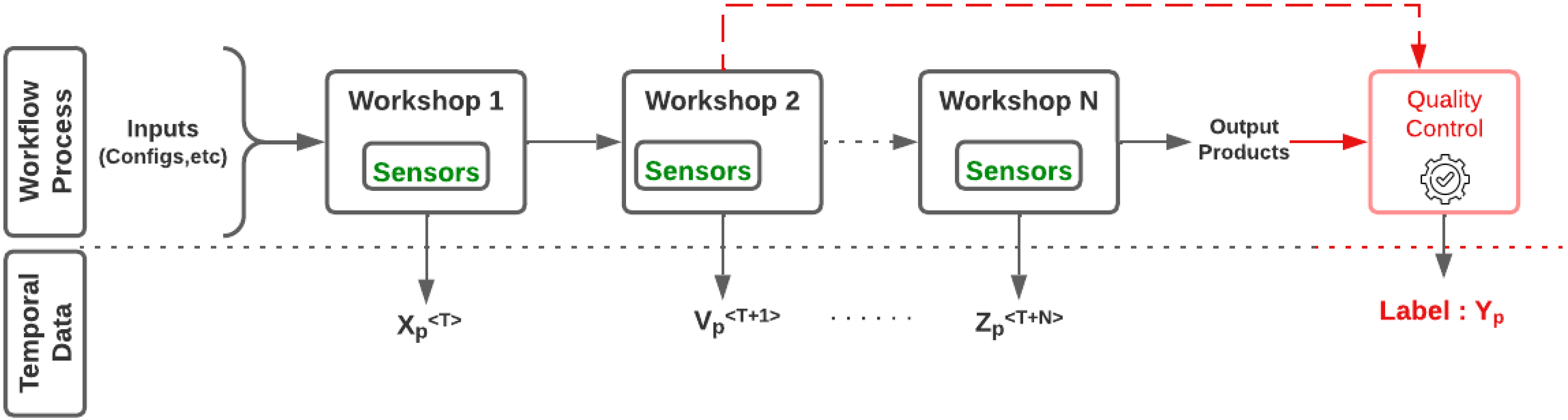
We are interested in classification problem of **weight control** using data collected in **mixing phase (Bloc 2)**



Use Case – Quality Control in Cake Factory

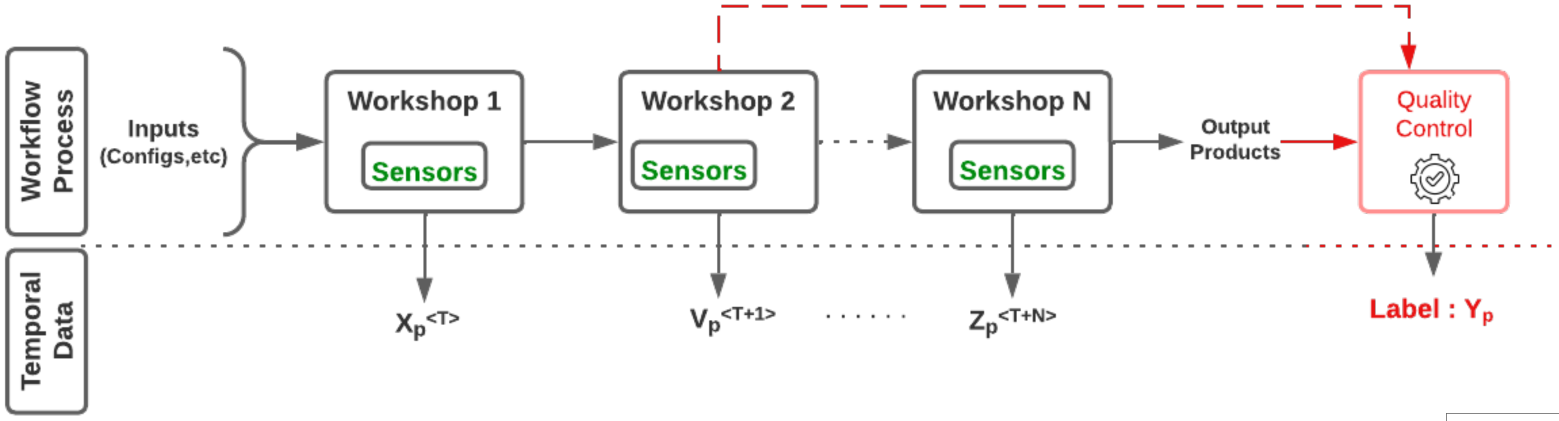
Example of Industrial Use Case

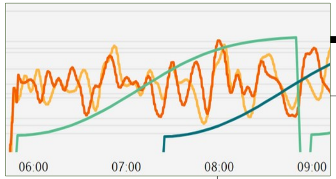
Quality Control / Predictive Maintenance



Use Case – Quality Control in Cake Factory

Workflow in Production line – Multi-Level Diagram





Time series data

DATETIME	M221_TS_01	M221_TP_01
2021-12-01 05:30:30	110.000000	6109.781280
2021-12-01 05:31:00	108.826667	6141.733192
2021-12-01 05:31:30	110.000000	6184.072883
2021-12-01 05:32:00	111.173333	6235.904952
2021-12-01 05:32:30	110.000000	6292.483363

Batch information

	ID_LOT_PATE	ETAPE	TIME_START	TIME_STOP
0	20211201-01	Mélange	2021-12-01 05:30:00	2021-12-01 05:44:10
1	20211201-01	Pétrissage	2021-12-01 05:44:10	2021-12-01 05:51:30
~~2~~	~~20211201-01~~	Repos	~~2021-12-01 05:50:00~~	~~2021-12-01 08:55:03~~
3	20211201-02	Mélange	2021-12-01 07:06:10	2021-12-01 07:20:20
4	20211201-02	Pétrissage	2021-12-01 07:20:20	2021-12-01 07:27:40

Train ML Model for Quality Control QICake

Time Series data

DATETIME	M221_TS_01	M221_TP_01
2021-12-01 05:30:30	110.000000	6109.781280
2021-12-01 05:31:00	108.826667	6141.733192
2021-12-01 05:31:30	110.000000	6184.072883
2021-12-01 05:32:00	111.173333	6235.904952
2021-12-01 05:32:30	110.000000	6292.483363

Batch information

	ID_LOT_PATE	ETAPE	TIME_START	TIME_STOP
0	20211201-01	Mélange	2021-12-01 05:30:00	2021-12-01 05:44:10
1	20211201-01	Pétrissage	2021-12-01 05:44:10	2021-12-01 05:51:30
~~2~~	~~20211201-01~~	Repos	~~2021-12-01 05:51:30~~	~~2021-12-01 06:55:02~~
3	20211201-02	Mélange	2021-12-01 07:06:10	2021-12-01 07:20:20
4	20211201-02	Pétrissage	2021-12-01 07:20:20	2021-12-01 07:27:40

Merge Data : Batch Identifier with the temporal data (INPUT DATA)

DATETIME	M221_TS_01	M221_TP_01	ID_LOT_PATE	ETAPE
2021-12-01 05:30:30	110.000000	6109.781280	20211201-01	Mélange
2021-12-01 05:31:00	108.826667	6141.733192	20211201-01	Mélange
2021-12-01 05:31:30	110.000000	6184.072883	20211201-01	Mélange
2021-12-01 05:32:00	111.173333	6235.904952	20211201-01	Mélange
2021-12-01 05:32:30	110.000000	6292.483363	20211201-01	Mélange
...	...	...	...	...
2022-03-01 12:13:30	118.854275	8742.818066	20220301-05	Pétrissage
2022-03-01 12:14:00	119.876393	8765.277318	20220301-05	Pétrissage
2022-03-01 12:14:30	120.888755	8787.267457	20220301-05	Pétrissage

Wight Quality Control per ID_PATE (TARGET DATA)

20211201-01	1
20211201-01	1
20211201-01	0
20211201-01	1
20211201-02	0
⋮ Only Zeros	
20211201-02	0
20211201-02	0

ID_LOT_PATE	QC_wight_err
20211201-01	1
20211201-02	0
20211201-03	0
20211201-04	1
20211201-05	0
20211202-01	1
20211202-02	0
20211202-03	1
20211202-04	0
20211202-05	0

Train ML Model for Quality Control - OlCake

Model Training & Evaluation

(1) Gradient Boosting Classifier

(~ with 50 estimators)

in Precision : 97.2% Train Recall: 96%

Precision : 96.9% Test Recall : 83.1 %

(2) Deep Neural Network

Train Precision : 83.3% Train Recall: 40.8 %

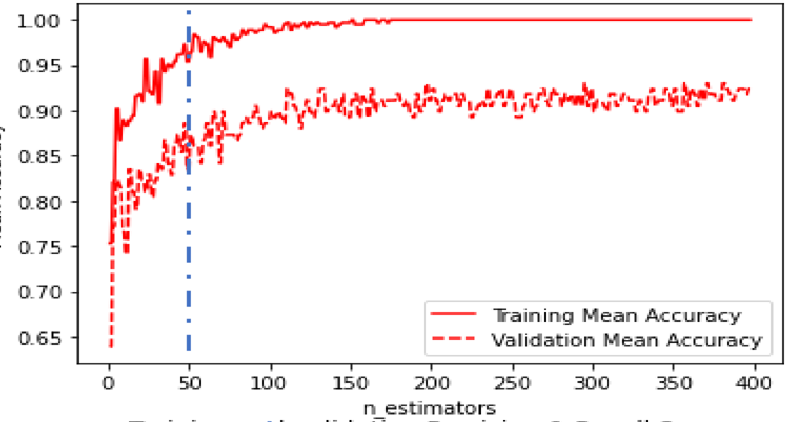
Test Precision : 82.9% Test Recall : 43%

(3) LSTM

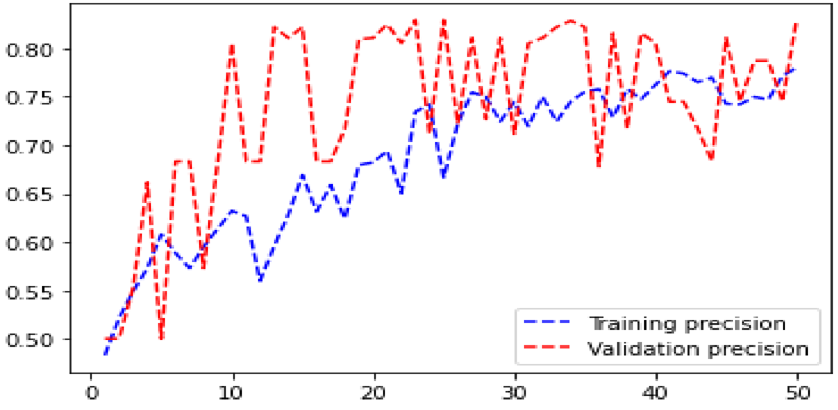
Train Precision: 67.1% Train Recall: 49.5%

Test Precision : 76.9% Test

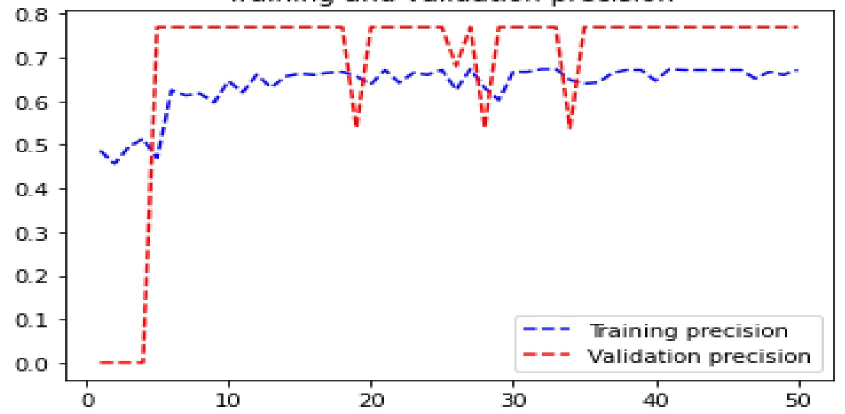
Training and validation Mean Accuracy



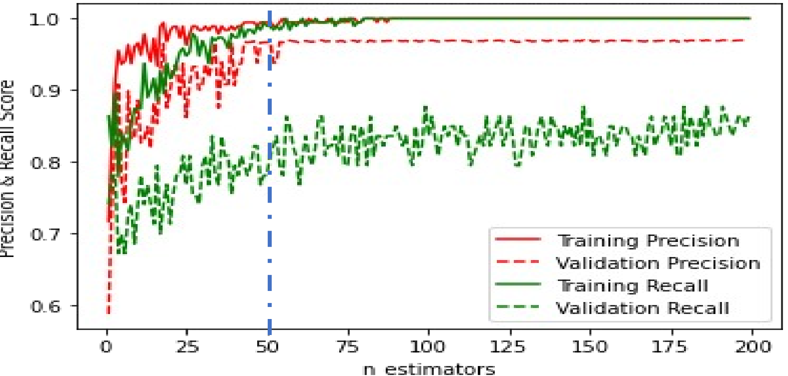
Training and validation precision



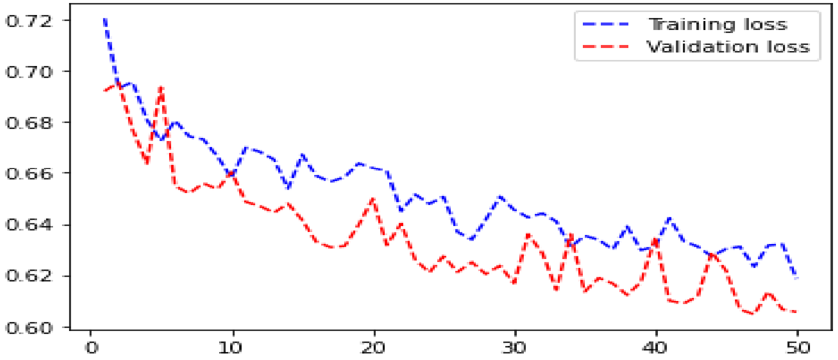
Training and validation precision



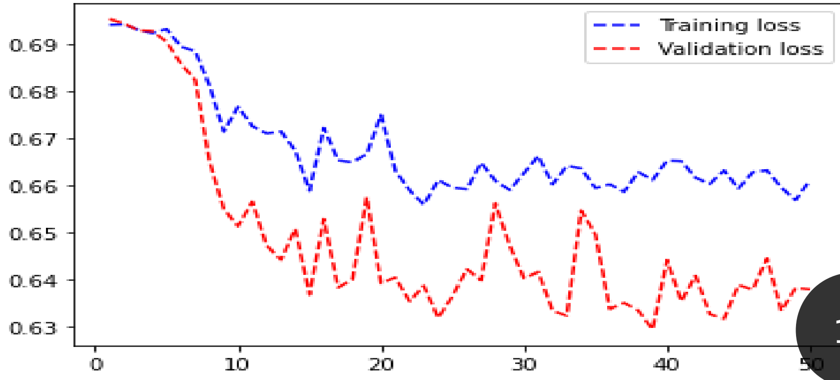
Training and validation Precision & Recall Score



Training and validation loss

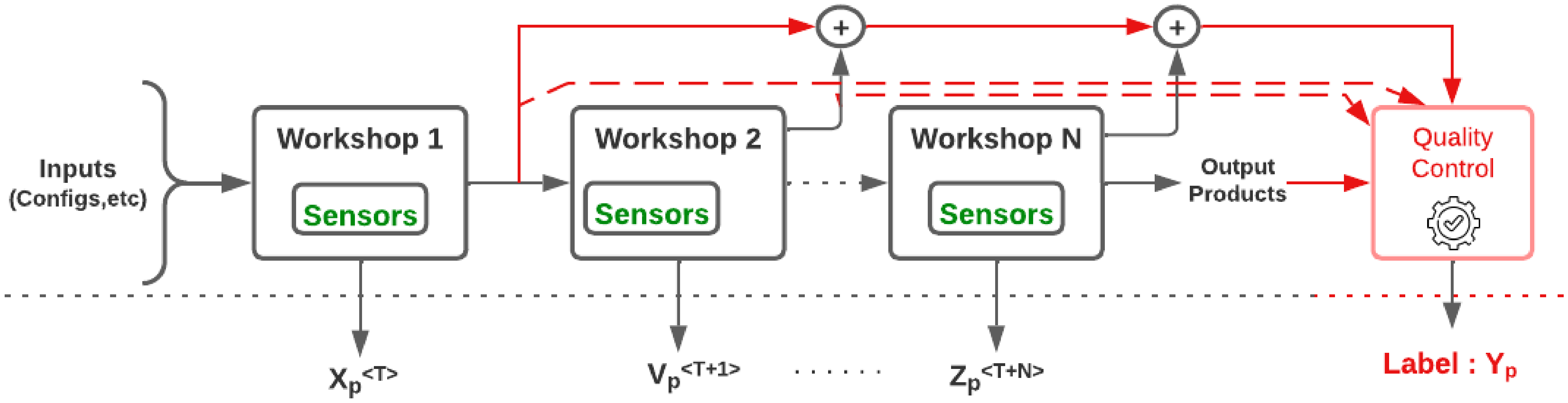


Training and validation loss



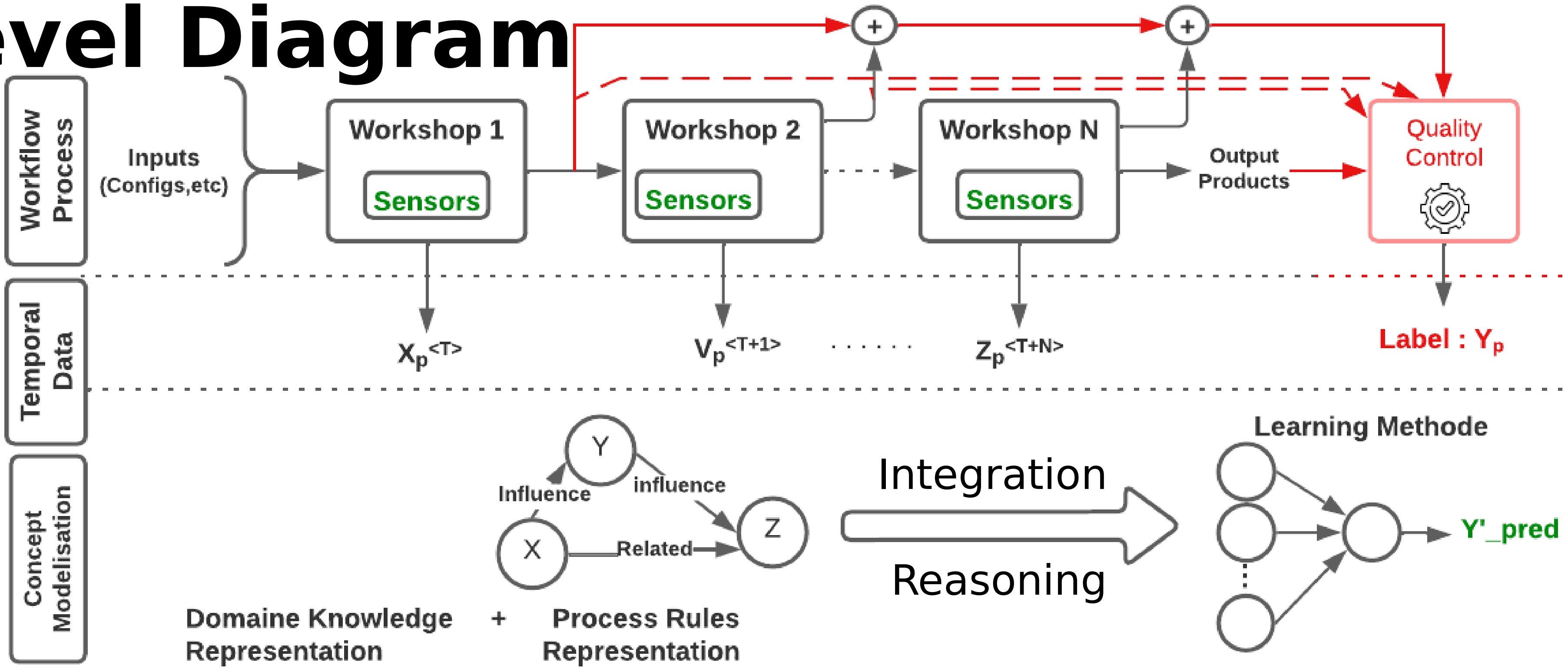
Example of Industrial Use Case

Quality Control / Predictive Maintenance



Use Case – Quality Control in Cake Factory

Workflow in Production line – Multi-Level Diagram



What is a Concept ?
What knowledge reasoning skill we need to integrate to our model ?

information about the process

Expert Knowledge

RESEARCH QUESTION

How to **Integrate** Domain-specific (industry 4.0) **knowledge** into **Machine Learning** to enhance its performance in downstream tasks ?

- **RQ1 : How** to **integrate numerical data** from **heterogeneous** observations to apply machine learning to downstream tasks? => **C1 [DATA INTEGRATION & HETEROGENEITY]**
- **RQ2 :** How to learn **the implicit knowledge** embedded in the industrial process and reason on it using machine learning (ML) models? => **C1 [DATA INTEGRATION & HETEROGENEITY] & C2 [Explicability]**

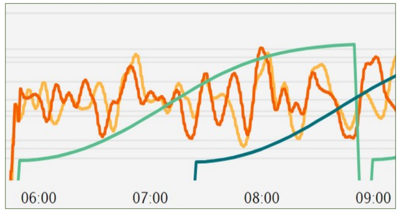
Problems & Challenges

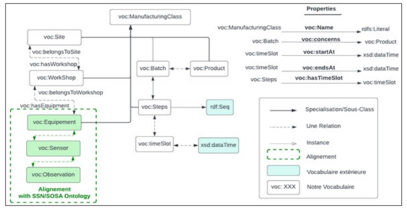
- **C1 [Data integration and heterogeneity] :** The overabundance and heterogeneity of available data **limits application of AI techniques** in the industry.
- **C2 [Explicability] :** ML/AI trained on raw data produces **black-box models** which lack **explicability**

Proposed Approach (Overview)

MINES Saint-Étienne COURBON Software

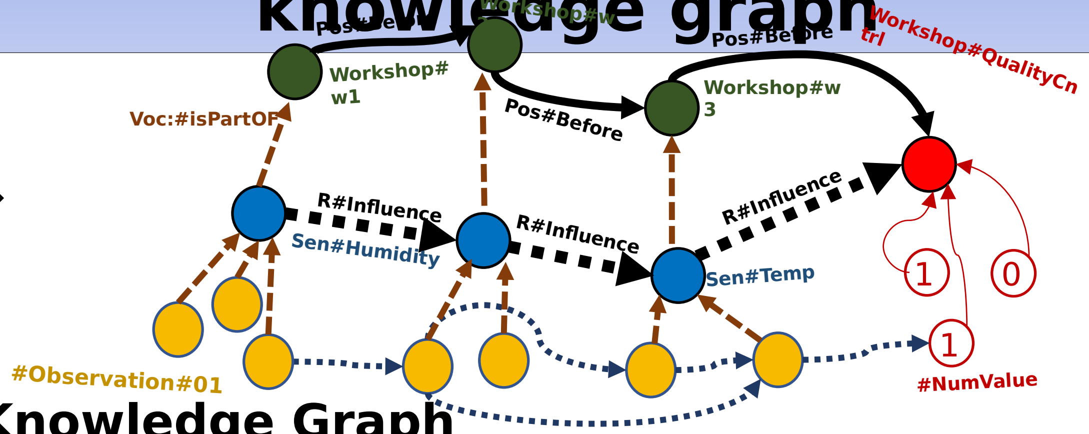
(1) Input Data





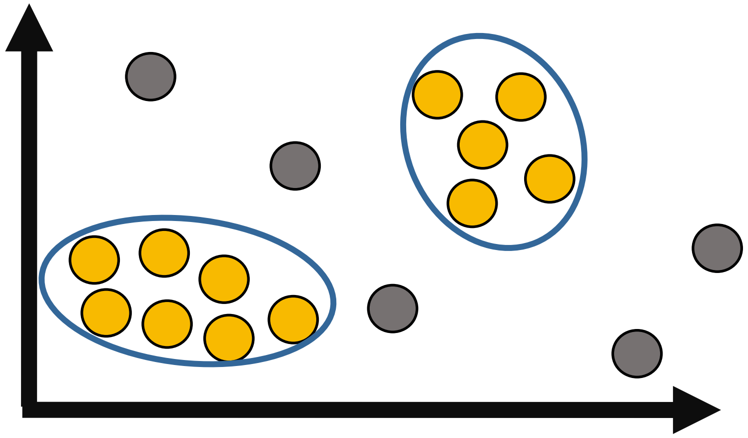
Process Data **Ontology**

(2) Integrate temporal data in knowledge graph



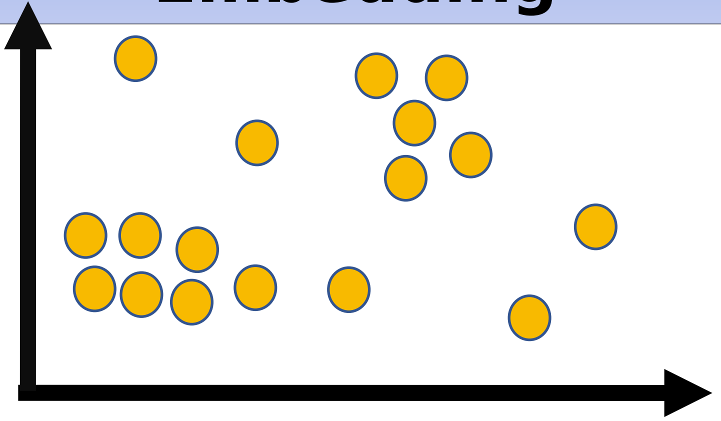
Knowledge Graph

(4) Downstream Task



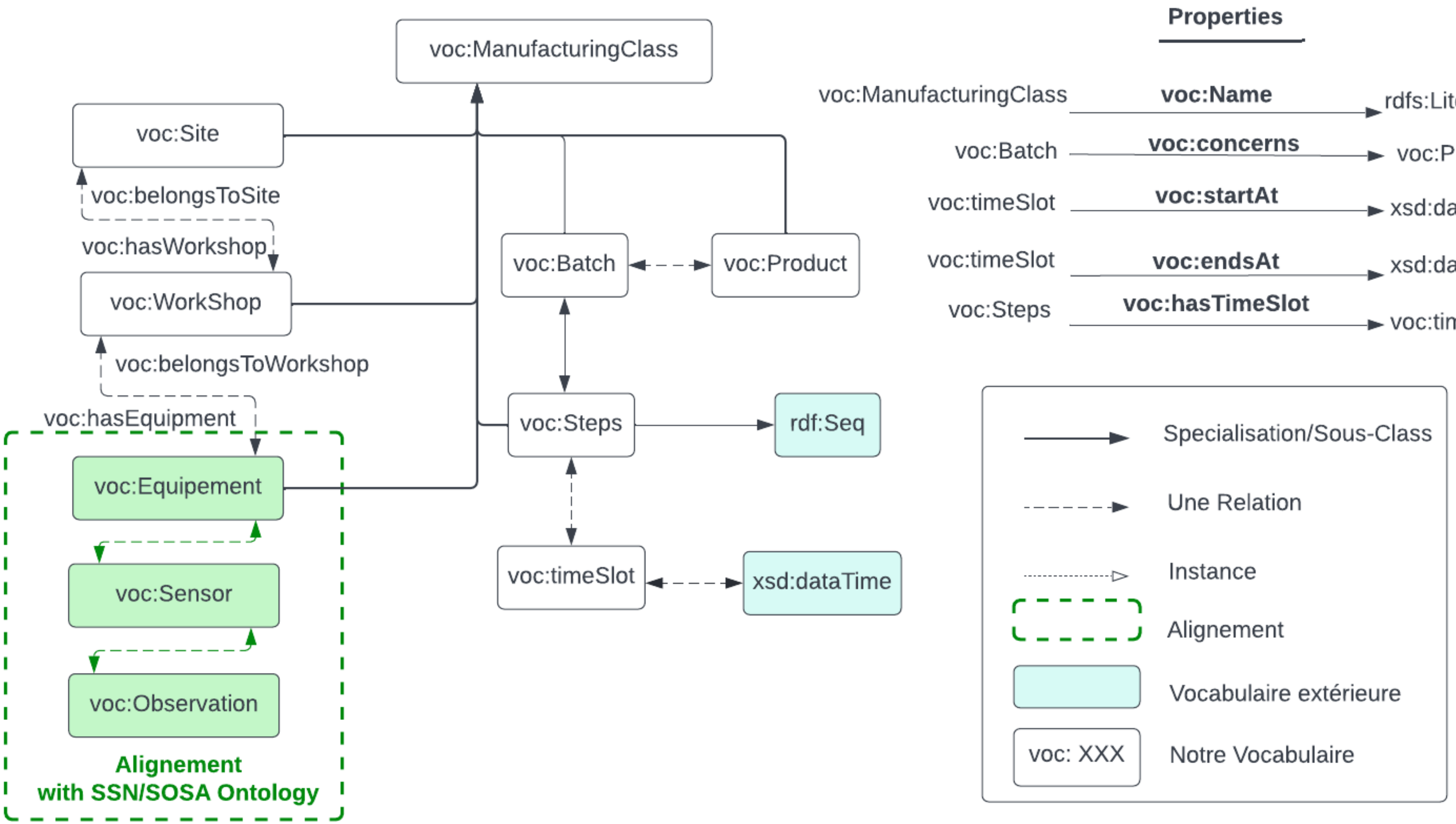
Classificatio
n

(3) Generate Knowledge Graph Embedding

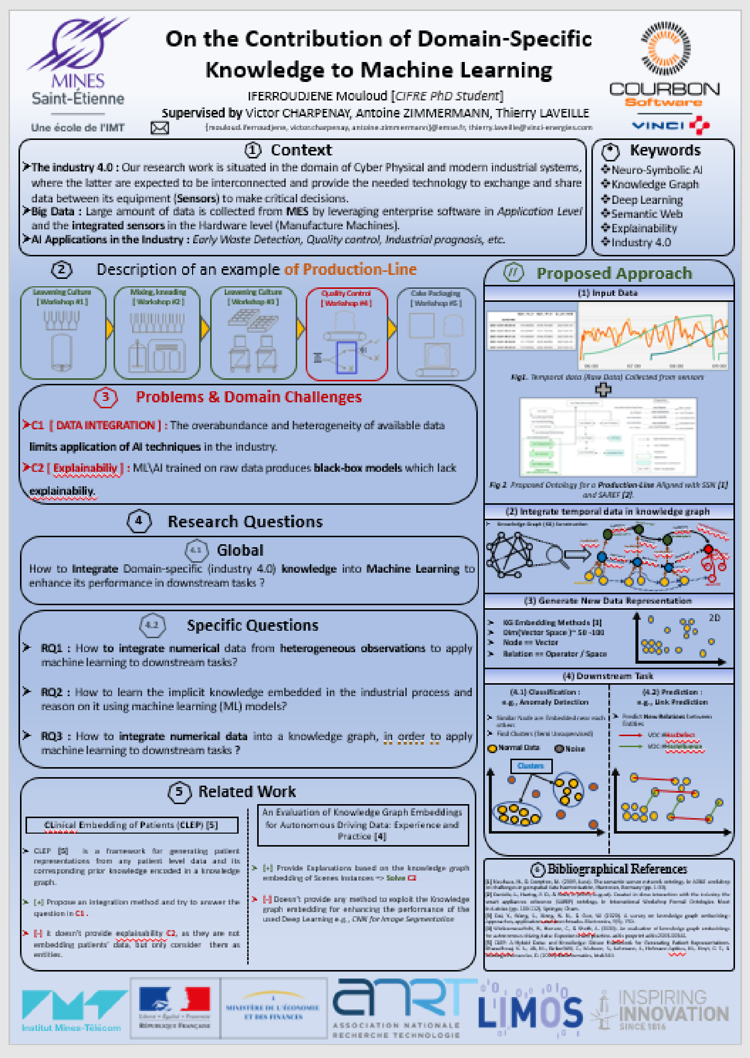


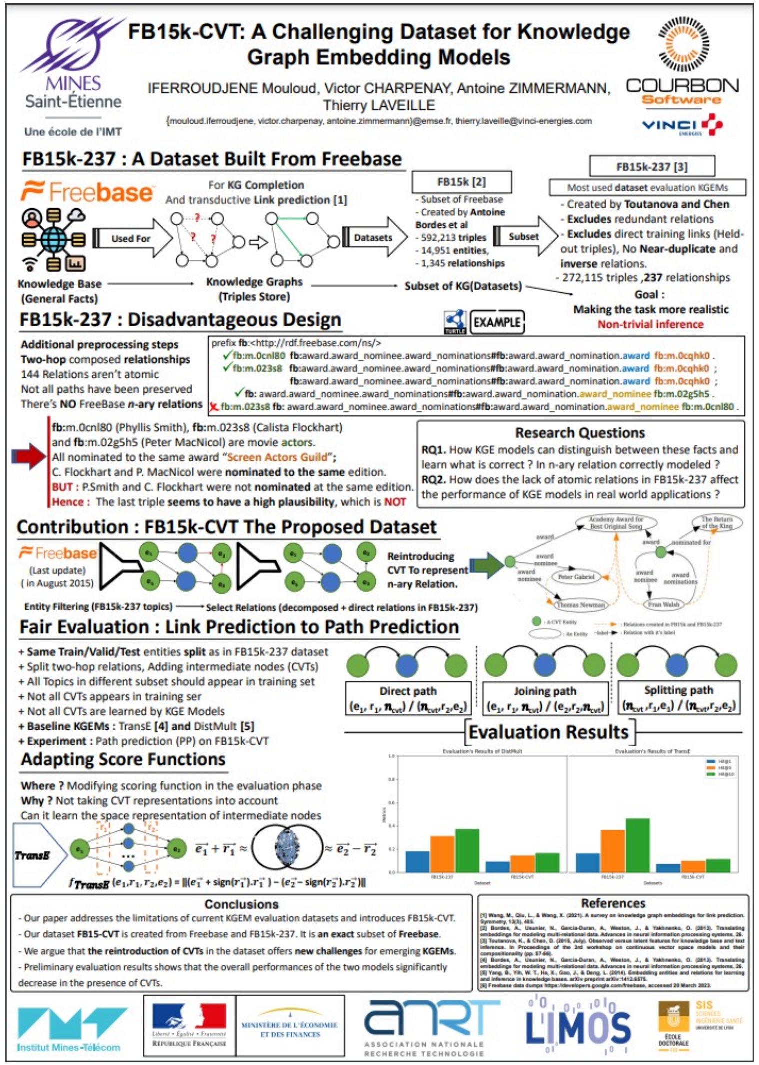
Industrial context - Modeling example

ed Hierarchical Ontology Proposed for the "OICake“ (cake factory) manufacturing pro



My Posters / Join Me To Check My WORK





FB15k-CVT: A Challenging Dataset for Knowledge Graph Embedding Models

Mouloud Iferroudjene[1,2], Victor Charpenay[1] and Antoine Zimmermann[1]

[1]Mines Saint-Etienne, Univ Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158 LIMOS, Saint-Etienne, France
[2]COURBON Software, 70 rue de la Montat – CS 60327, 42015 Saint-Etienne, France

Abstract
Knowledge Graphs (KGs) are an essential component of neuro-symbolic AI. KG Embedding Models (KGEMs) are used to represent elements of a KG (its entities and relations) in a vector space, to enable efficient processing and reasoning over knowledge. Most KGEMs are evaluated against datasets derived from the Freebase KG: FB15k and FB15k-237. In this paper, we identify limitations in these datasets with respect to Compound Value Types (CVTs), which are nodes introduced in Freebase as a substitute for n-ary relations. In FB15k and FB15k-237, CVTs have been removed, thereby eliminating valuable information. To evaluate whether KGEMs can learn semantically accurate representations of entities and relations in Freebase, we introduce here a new dataset named FB15k-CVT, which reintroduces the deleted CVT nodes. In a preliminary evaluation, we assess the limitations of baseline KGEMs (TransE, DistMult) in the presence of CVTs. The evaluation suggests that KGEMs based on tensor decomposition are more promising than translational models but, most of all, it calls for further experiments with KGEMs that can answer conjunctive queries or that preserve logical entailment.

Keywords
Knowledge Graphs, Neurosymbolic AI, Knowldge Graph Embeddings Models, FB15K-237

1. Introduction

Knowledge graphs (KGs) have become an essential component of neuro-symbolic AI research. A KG is a uniform source of information in which physical-world entities are represented as vertices of a directed edge-labeled graph. In the context of representation learning, edge labels of a KG are called relations, and its edges are called facts or triples [1].

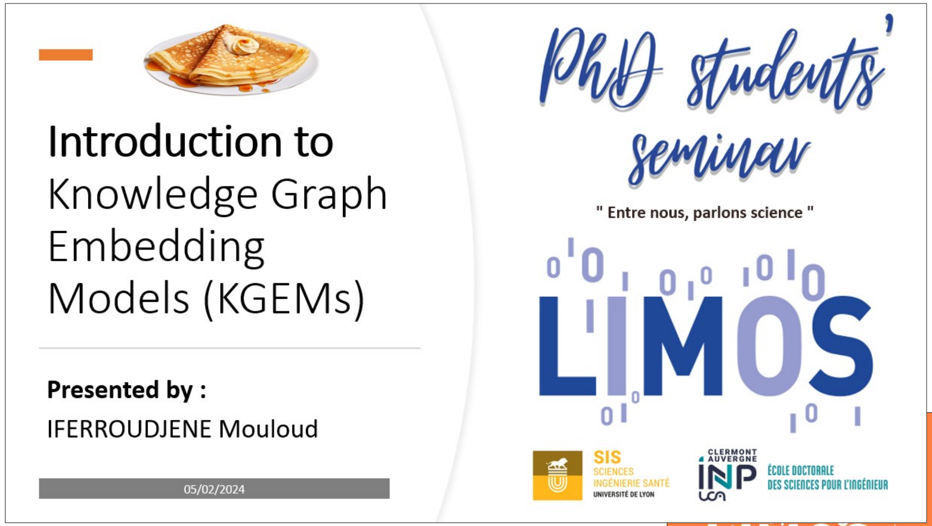
KGs can be leveraged in a great variety of AI applications. Over the past decade, many KG Embedding Models (KGEMs) have been developed for that purpose [1, Sec. 4.2]. By representing entities and relations as numeric structures in a vector space, KGEMs provide a way to integrate both symbolic and sub-symbolic knowledge, enabling efficient processing and reasoning over complex and heterogeneous data. Most KGEMs are evaluated against datasets that are derived from Freebase[1], a (now archived) public KG containing millions of entities and billions of facts.

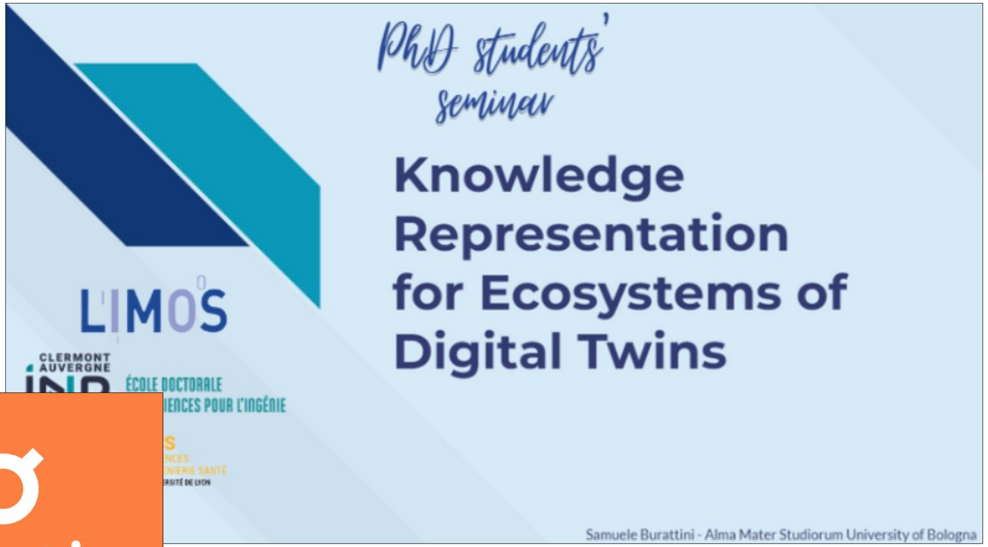
In KGEM research, the most notable datasets derived from Freebase are FB15k and FB15k-237. FB15k is a subset of Freebase that includes 15k entities selected among the most frequent entities

NeSy 2023, 17th International Workshop on Neural-Symbolic Learning and Reasoning, Certosa di Pontignano, Siena, Italy
mouloud.iferroudjene@emse.fr (M. Iferroudjene); victor.charpenay@emse.fr (V. Charpenay); antoine.zimmermann@emse.fr (A. Zimmermann)
© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)
[1]Freebase data dumps https://developers.google.com/freebase, accessed 20 March 2023.

Membres :
- Valentin GUIEN
- Mouloud IFERROUDJENE
- Frederic HAYEK
- Aurelien MOMBELLI

PhD students' seminar





The End

Thank You For Your Attention

Any Questions ?