ULCO Université du Littoral Côte d'Opale — HESTIM Engineering & Business School — ENSAM École Nationale Supérieure d'Arts et Métiers, Université Hassan II de Casablanca

5th International Conference on Big Data and Internet of Things (BDIoT'21)

Towards an Automatic Assistance framework for the selection and configuration of machine-learning based data analytics solutions in industry 4.0

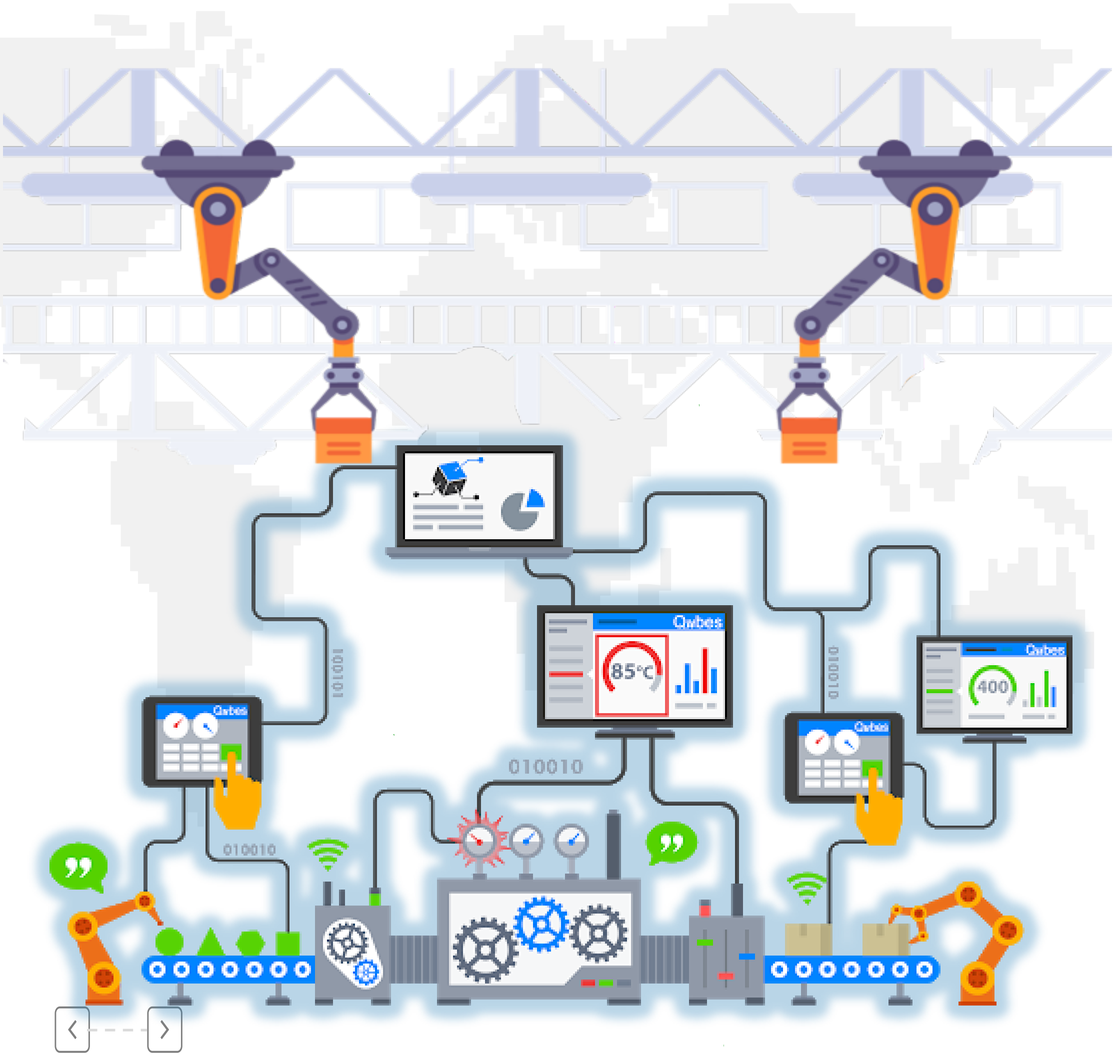
Moncef Garouani[1,2,3] , Mohamed Hamlich[2] , Adeel Ahmad[1], Mourad Bouneffa[1], Arnaud Lewandowski[1] and Gregory Bourguin[1]

[1] LISIC Laboratory, ULCO University, Calais, France
[2] ISSIEE Laboratory, ENSAM, University of Hassan II, Casablanca, Morocco
[3] Study and Research Center for Engineering and Management, HESTIM, Casablanca, Morocco

PLAN

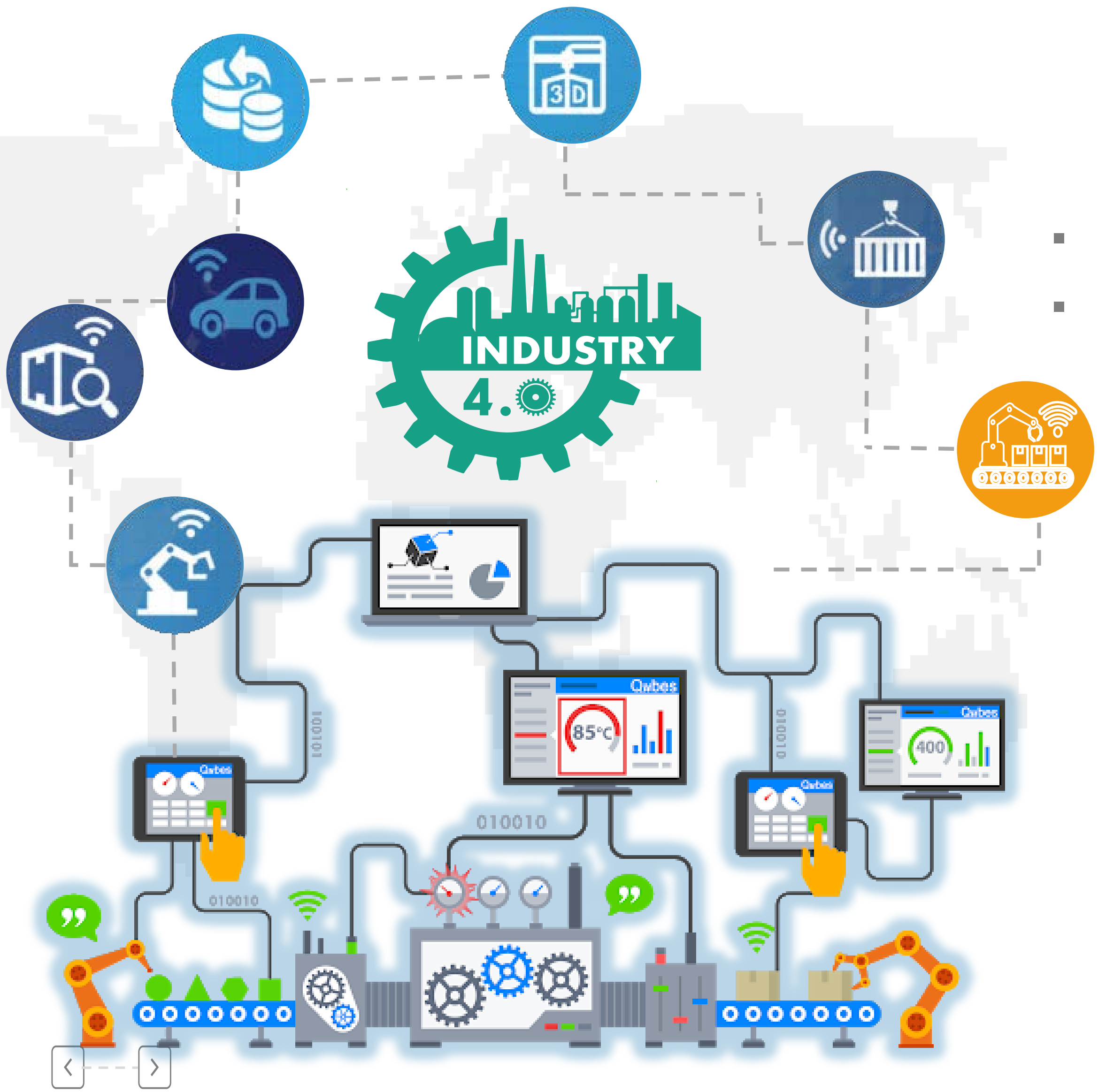
- Introduction
- Meta-learning
- Background and literature review
- Proposed system
- Experiments and Results
- Conclusion & Perspectives

Context (1/2)



- Automation and data exchange in manufacturing industry
- Smart factory
- Data analytics approaches in classical manufacturing factories
- Recent development trends vs Industrial Data Analytics (IDA)

Context (2/2)

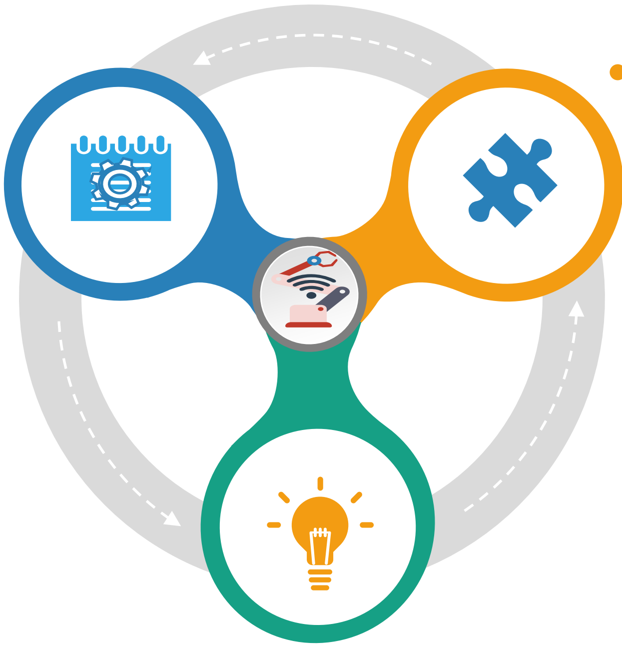


- **Industrial data analytics in Industry 4.0**
- **Industry 4.0 vs Machine Learning**
  - Industry Internet of Things (IIoT)
  - Cloud Computing and Edge data computation
  - production systems in smart factories
  - Interconnected continuous data exchange
  - Data lakes
  - Industrial Big Data Analysis

Motivation

Industrial Data Analytics

- **Industrial Data Analytics *(IDA)*** provides assistance to engineers in Industrial decision-making.
- These systems provide recommendations in finding the right diagnosis or the optimal decision.



Industrial Data Science

❑ **Challenges in building ML predictive models with big manufacturing data**

- Efficiently selecting ML algorithms
- Efficiently configuring related hyperparameters

Objectives

- Autonomous industrial smart applications
- Rapid collaboration among industrial actors and data scientists
- Decision-support systems
- Explainability of intelligent predictions

Introduction

Industry 4.0 vs Machine learning

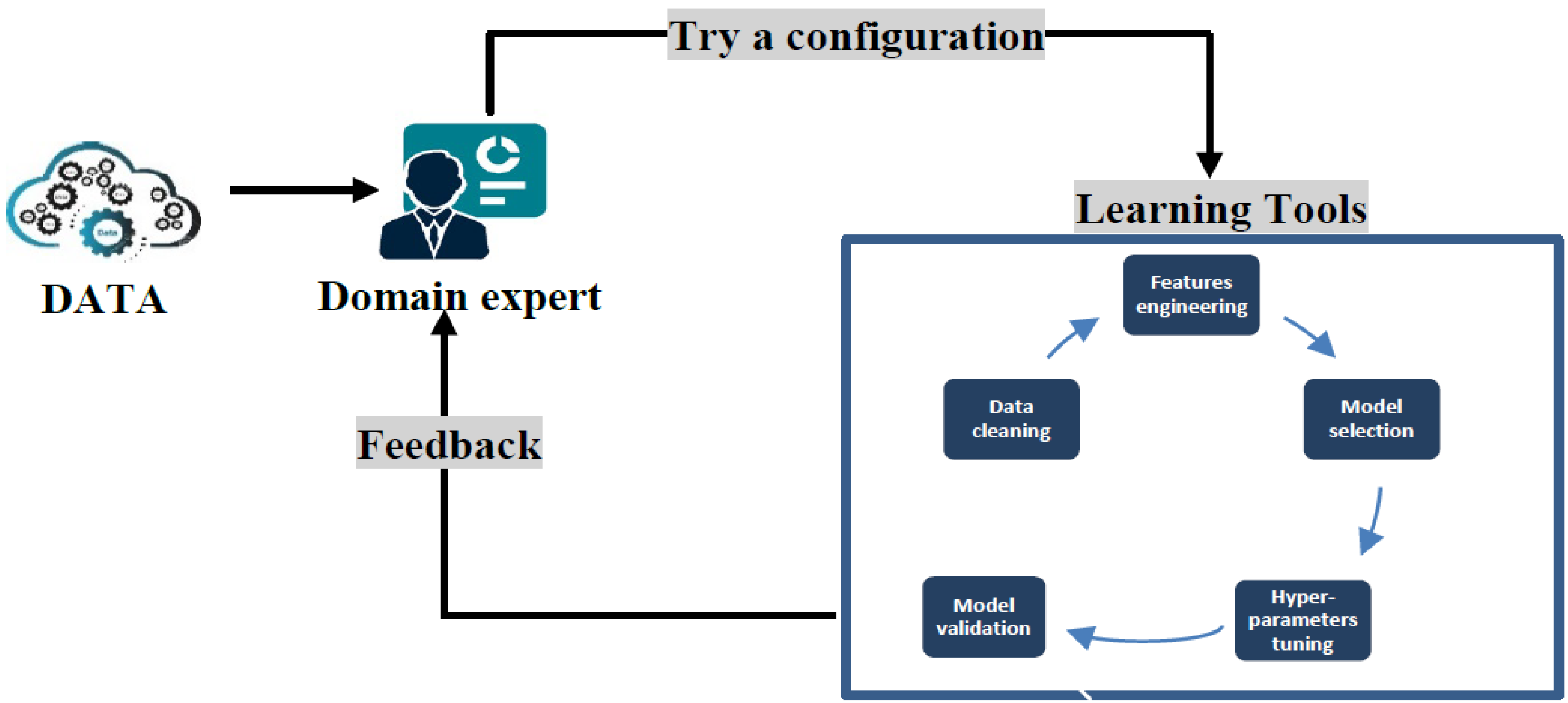
- Predictive analytics
- Data-driven decision making
- Interpretation of data patterns
- Make value from massive industrial data

IDA applications

- Process level
- Machine level
- Shop floor level
- Supply chain level

Introduction

Human tuning process configuration



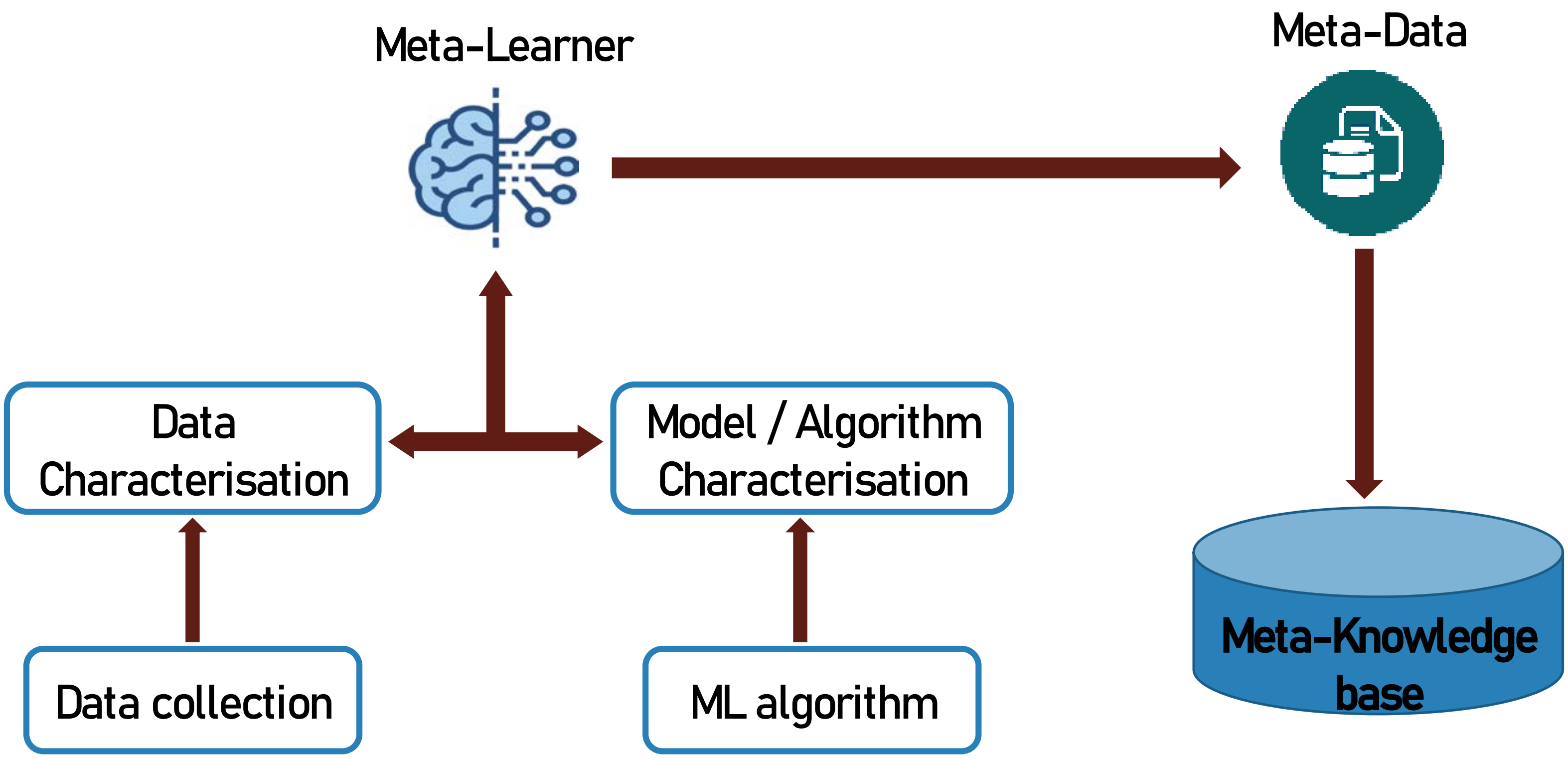
Meta learning (1/2)

- *Learn to learn* using previous knowledge from related tasks.

- Algorithm Learning (selection)
  - Select learning algorithms according to the characteristics of the instance.
- Hyper-parameter Optimization
  - Select hyper-parameters for learning algorithms.

- "Algorithms show similar performance for the same configuration for similar problems"

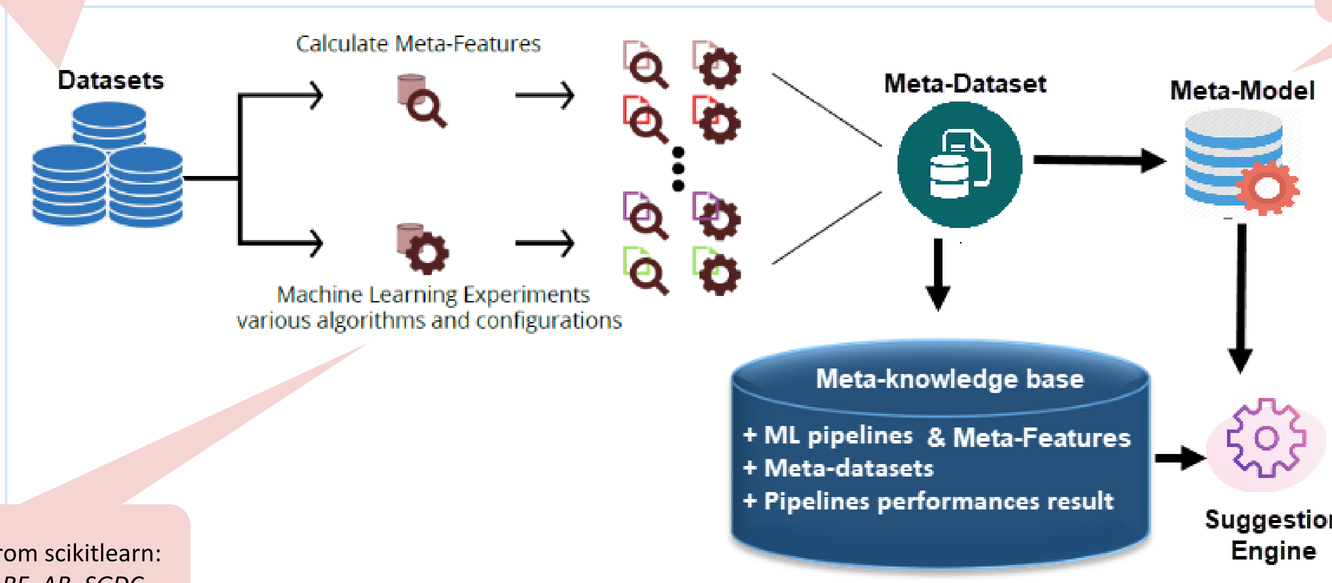
Meta learning (2/2)



I- Learning phase

200 real-world manufacturing classification datasets

k-Nearest Neighbor (KNN)



8 ML algorithms from scikitlearn: *SVM, LR, DT, ET, RF, AB, SGDC.*

Figure 1: The workflow of the learning phase

II- Recommendation phase

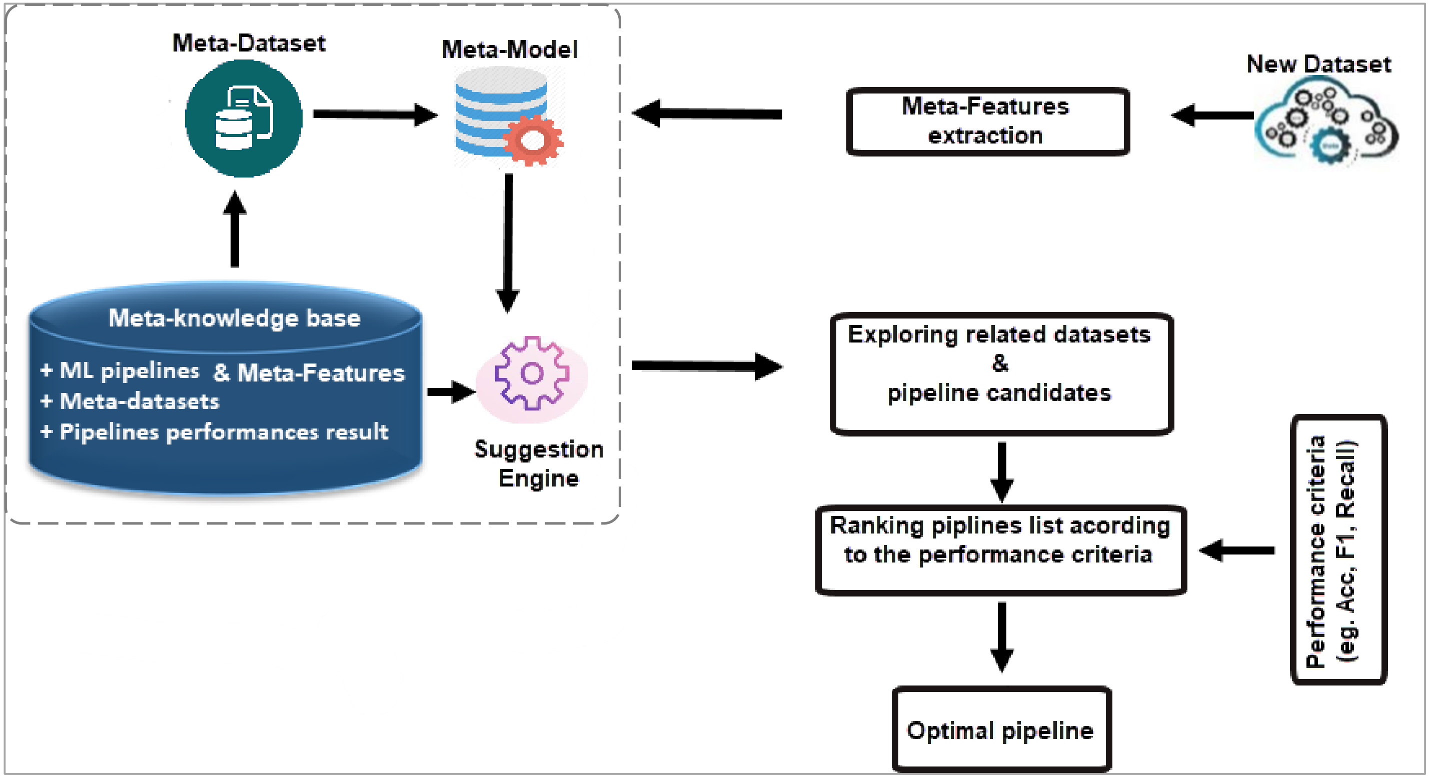


Figure 2: The workflow of the recommendation phase

Framework

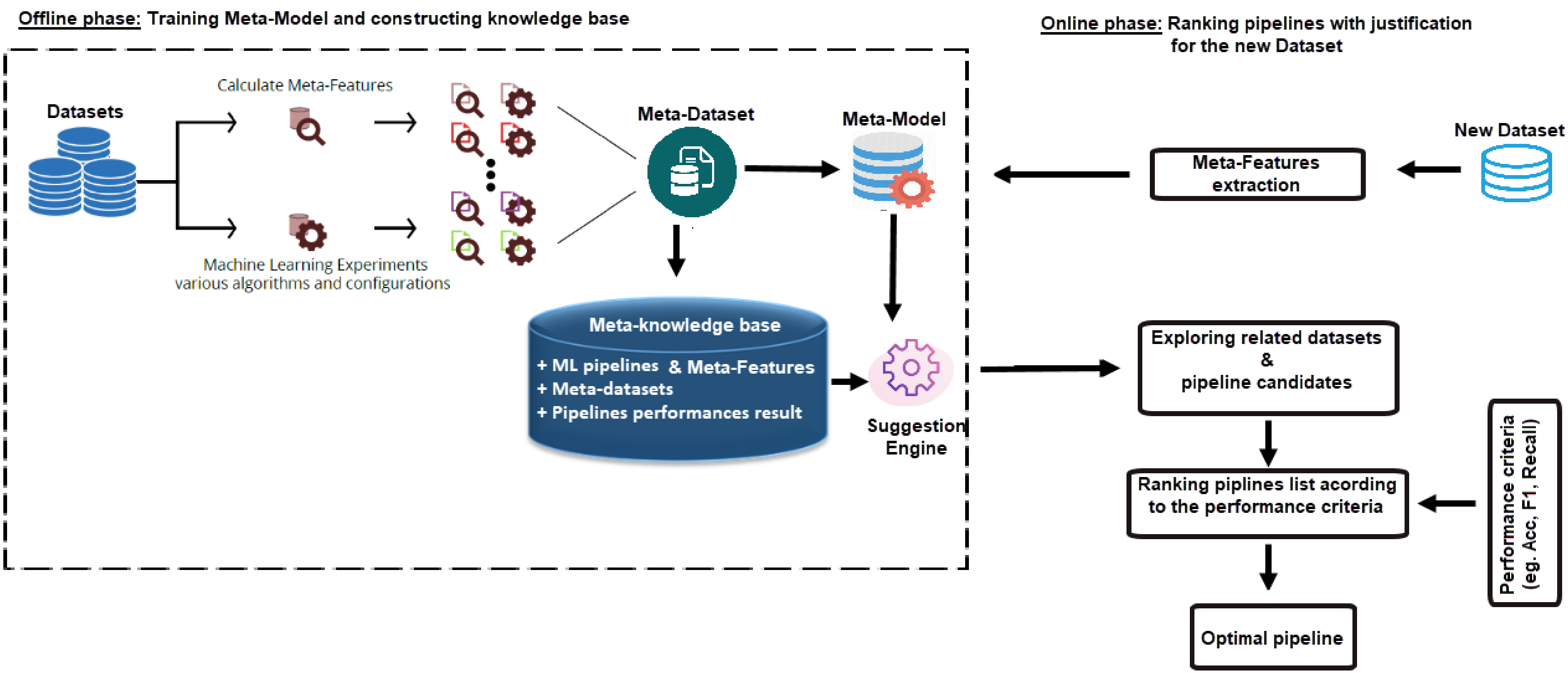


Figure 3: The workflow of the proposed framework

Analysis evaluation

Evaluation Benchmark

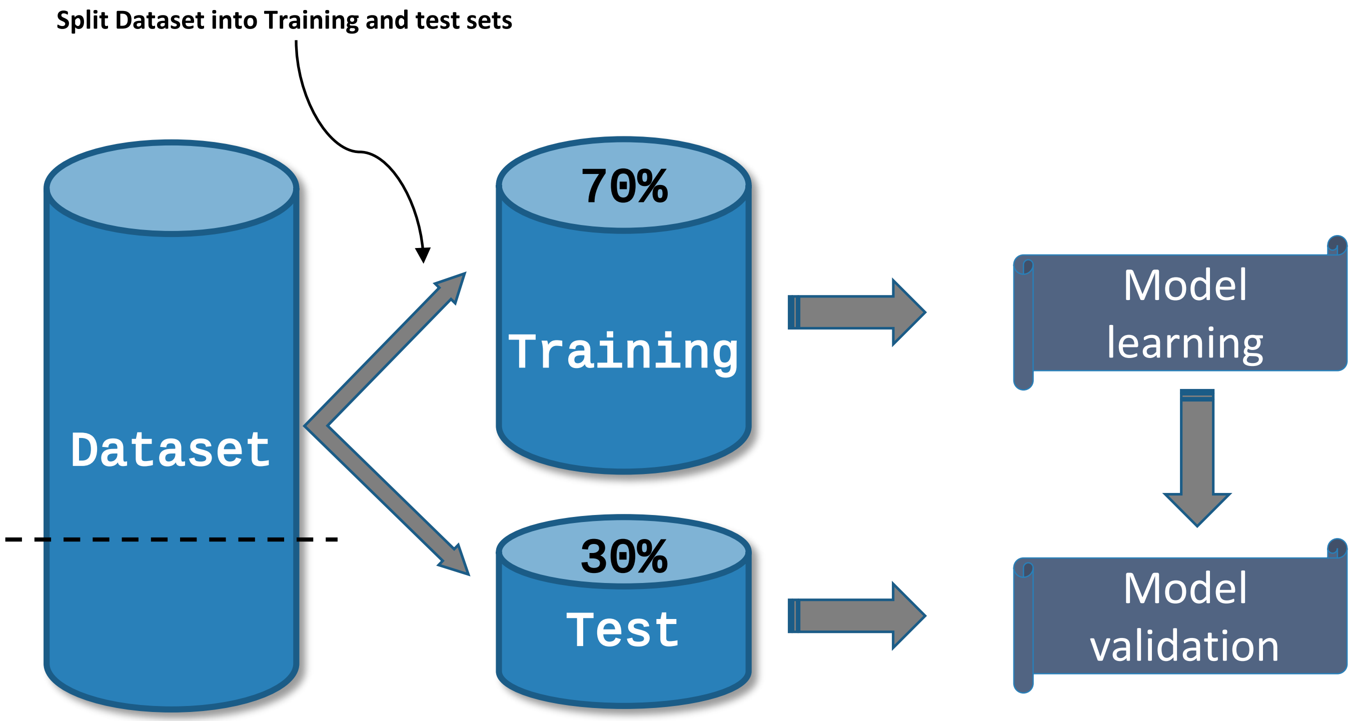
Dataset	Num. of Classes	Num. of Instances	Task
(Mazumder et al.,2021)	4	959	Failure risk analysis of pipeline networks
(Benkedjouh et al., 2015)	2	61000	RUL prediction
(Saravanamurugan et al., 2017)	3	2000	Chatter prediction
(Costa and Nascimento, 2016)	2	60000	APS system failure prediction
(Baldi et al., 2014)	2	98050	high-energy physics data analyses
(Tian et al., 2015)	7	1941	Faults detection

Table 1 : List (sample) of Datasets used in the evaluation

- Mazumder, Ram K., Abdullahi M. Salman, and Yue Li. "Failure risk analysis of pipelines using data-driven machine learning algorithms." SF 89 (2021): 102047.
- Benkedjouh, Tarak, et al. "Health assessment and life prediction of cutting tools based on support vector regression." JIM, 26.2 (2015): 213-223.
- Saravanamurugan, S., et al. "Chatter prediction in boring process using machine learning technique." IJMR, 12.4 (2017): 405-422.
- Costa, Camila Ferreira, and Mario A. Nascimento. "Ida 2016 industrial challenge: Using machine learning for predicting failures." IDA. Springer, Cham, 2016
- Baldi, Pierre, Peter Sadowski, and Daniel Whiteson. "Searching for exotic particles in high-energy physics with deep learning." NC, 5.1 (2014): 1-9.
- Tian, Yang, Mengyu Fu, and Fang Wu. "Steel plates fault diagnosis on the basis of support vector machines." Neurocomputing 151 (2015): 296-303.

Analysis evaluation

Evaluation strategy



Analysis evaluation

Evaluation results

Dataset	Recommended config. result	Paper result	Pipeline with default config.
(Mazumder et al.,2021)	**93.74**	85	80.24
(Benkedjouh et al., 2015)	**99.41**	98.95	93.88
(Saravanamurugan et al., 2017)	**97.06**	95	86.12
(Costa and Nascimento, 2016)	**99.10**	92.56	92.34
(Baldi et al., 2014)	85.59	88	69.45
(Tian et al., 2015)	**99.54**	80.74	76.23

Table 2: Performances of the proposed framework

Analysis evaluation

Evaluation results

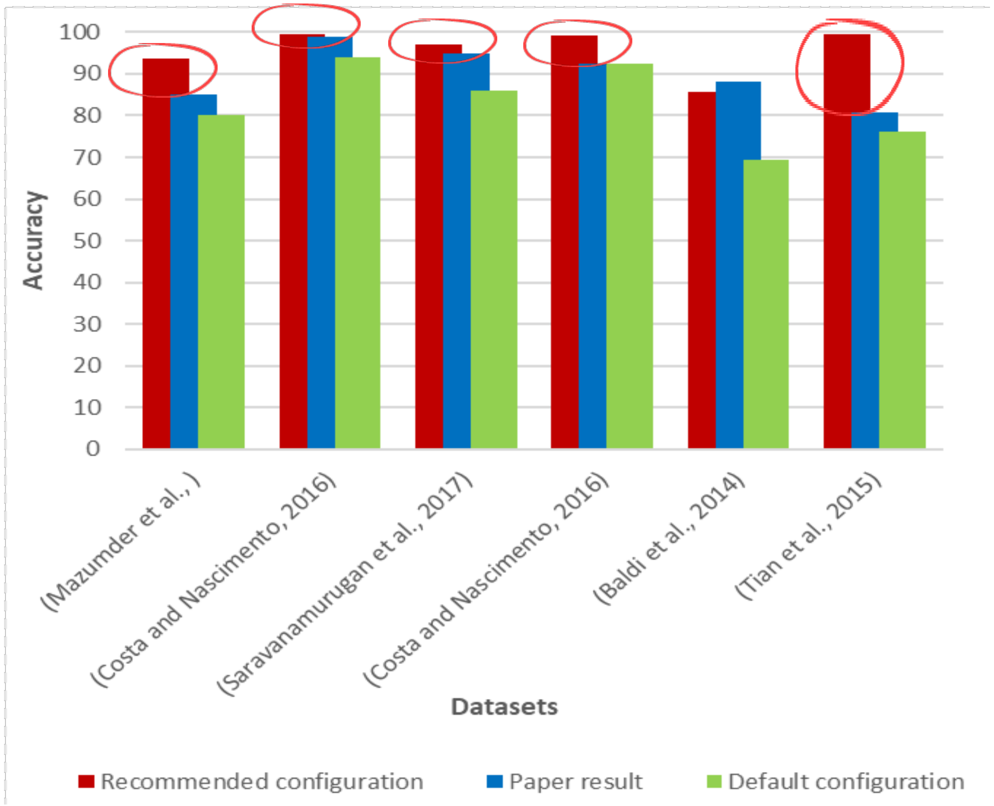


Figure 4: Comparative results of the effectiveness of AutoML over default classic ML configuration and domain expert (industrial researchers) configurations.

Conclusion

This work studied the effectiveness of AutoML techniques for the manufacturing related problems.

We proposed the design of an AutoML based decision support systems for Industry 4.0 actors and researchers.

We have implemented the proposed Framework using **8** ML algorithms on **200** real-world dataset from different industry 4.0 levels.

We evaluated the proposed system on a benchmark of **20** binary and multi-class classification problems from different industry 4.0 levels.

Our system achieves convincing results.

The obtained results are more accurate than the results from the datasets related papers.

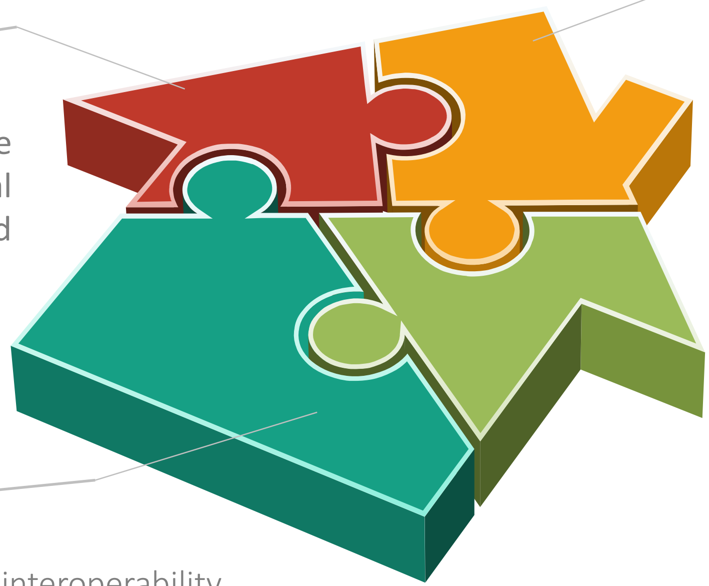
The comparative analysis reveals in the majority of the cases that the recommended configurations yield better performances,

Perspectives

The next planned steps include:

1. Further validation of the proposed framework in other real world applications with a larger and more diverse problems .

2. Add support for further data formats and ML algorithms



3- Work on the explainability and interoperability aspect of AutoML systems as being black-boxes.

THANK YOU FOR YOUR ATTENTION

ulco UNIVERSITÉ DU LITTORAL CÔTE D'OPALE

hestim Engineering & Business School

ENSAM ÉCOLE NATIONALE SUPÉRIEURE D'ARTS ET MÉTIERS UNIVERSITÉ HASSAN II DE CASABLANCA

Toward an Automatic Assistance framework for the selection and configuration of machine-learning based data analytics solutions in industry 4.0