Francesco Vitale

Combining Process Mining and Machine Learning for Anomaly Detection

Tutor: Prof. N. Mazzocca
co-Tutor: Dr. F. Papa

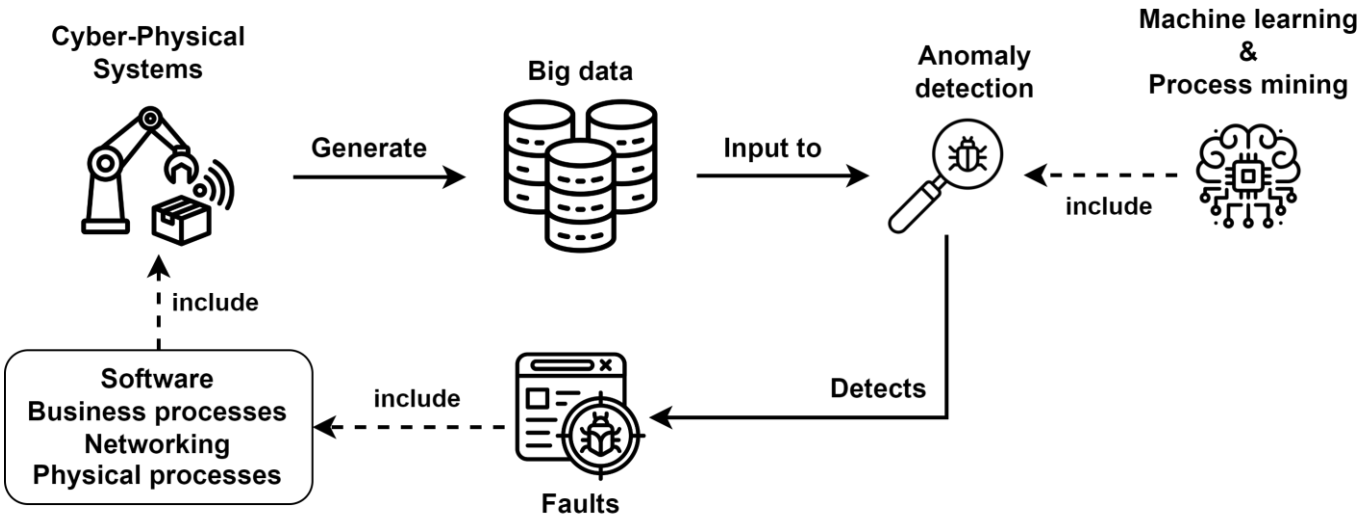
Cycle: XXXVII
Year: 2021/2022

My background

- BSc and MSc in Computer Engineering @ **University of Naples Federico II**
- PhD start date: 01/11/21 @ **SECLab research group**
- Company-funded scholarship with **Hitachi Rail S.T.S. @ Automation Value Added Services unit**[P1]
- In-progress collaborations with several Italian and foreign universities
  - **University of Messina** (UNIME)[J3]
  - **University Campus Bio-Medico of Rome** (UCBM)[J4]
  - **RWTH Aachen University** (RWTH)[J5]
  - **Linnaeus University** (LNU)[J6]
- **Period abroad** at the **Process and Data Science (PADS) group @ RWTH**
  - Supervisor: Prof. Dr. Ir. W. M. P. van der Aalst
  - Start date: 01/02/23
  - End date: 31/07/23

Research field of interest

- **Anomaly Detection (AD) with Machine Learning (ML) and Process Mining (PM) in Cyber-Physical Systems (CPSs)**



Summary of study activities

- **Ad-hoc PhD courses**
  - IoT Data Analysis
  - Using Deep Learning properly
  - Cambridge Advanced C1

- **Study period abroad at the PADS group @ RWTH**
  - The group is specialized in PM algorithms, methodologies and employment in application domains

- **Visited the COSERITY lab @ UCBM**
  - Held a seminar and started new collaborations

Research activity: The problem

- **AD in CPSs requires addressing several layers of complexity**
  - Software, business processes, networking, physical processes
- **There is no one-size-fits-all solution**
  - The blind application of ML and PM algorithms is not sufficient

ML is suited for numerical data, e.g., time series

- Clustering, dimensionality reduction, neural networks, etc.

PM handles event logs, e.g., business process activities

- Process discovery and conformance checking

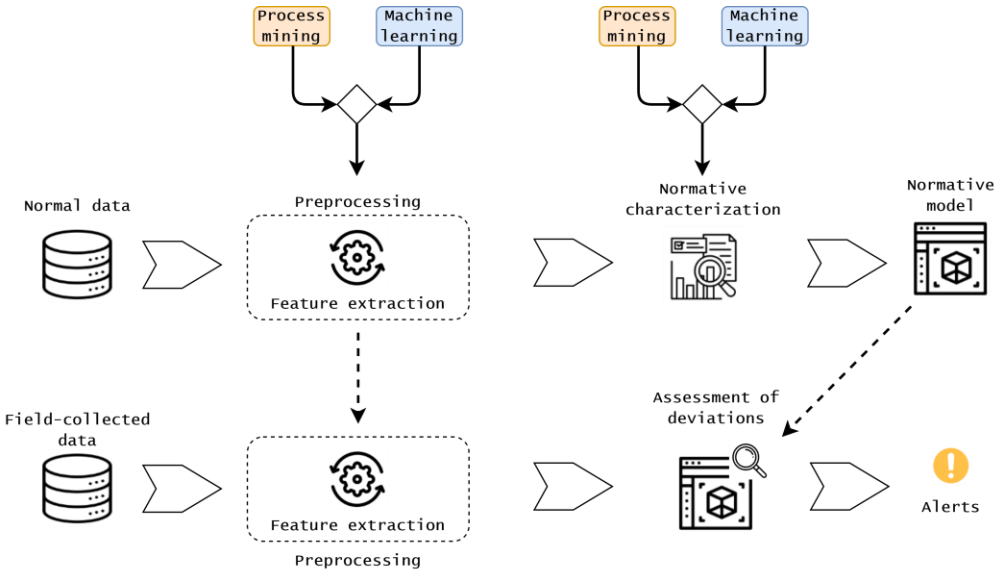
- **The strengths of ML and PM can be combined for AD in CPSs**

Research activity: The objective and methodology

Development of an overarching framework for AD in CPSs with ML and PM

Methodology:

- Address **research gaps on AD in CPSs with PM by integrating ML**;
- **Compare results** with similar approaches **and assess strengths and weaknesses**;
- **Formalize the steps of the framework for specific data and domains.**



- **Preprocessing by feature extraction**
- **Normative characterization**
- **Assessment of deviations**

Future work

- **Future work involves finalizing my contributions, formalizing the overarching framework, and developing new research directions, including:**

1. **Modeling and simulating CPSs by PM for implementing their Digital Twin and discover cyber attacks**

2. **Automatic response to anomalies for recovery towards acceptable conditions**

Products

[J1]	A. De Benedictis, F. Flammini, N. Mazzocca, A. Somma, F. Vitale, "*Digital Twins for Anomaly Detection in the Industrial Internet of Things: Conceptual Architecture and Proof-of-Concept*," IEEE Transactions on Industrial Informatics, vol. 19, no. 12, pp. 11553-11563, 2023, https://doi.org/10.1109/TII.2023.3246983
[J2]	M. Cinque, L. De Simone, N. Mazzocca, D. Ottaviano e F. Vitale, "*Evaluating Virtualization for Fog Monitoring of Real-time Applications in Mixed-Criticality Systems*," Real-Time Systems, 2023. (**accepted for publication**)
[J3]	F. Vitale, F. De Vita, D. Bruneo e N. Mazzocca, "A Process Mining-based Unsupervised Anomaly Detection Technique for the Industrial Internet of Things." (submitted to Internet of Things (Netherlands), **currently under minor review**).
[J4]	S. Guarino, F. Vitale, F. Flammini, L. Faramondi, N. Mazzocca e R. Setola, "A Two-Level Fusion Framework for Cyber-Physical Anomaly Detection." (submitted to IEEE Transactions on Industrial Cyber-Physical Systems, **currently under major review**).

Products

[J5]	F. Vitale, M. Pegoraro, W. M. P. Van der Aalst e N. Mazzocca, “A Comparison Framework for Control-Flow Anomaly Detection in Event Logs of Information Systems.” (submitted to IEEE Transactions on Knowledge and Data Engineering, **currently waiting for first decision**).
[J6]	F. Vitale, F. Flammini, M. Caporuscio e N. Mazzocca, “Combining Process Mining and Machine Learning for Direct Monitoring in Resilient Computer Systems.” (**in-progress**)
[P1]	Prototype for anomaly detection in the height-stagger profile of trains' pantograph during their journeys (**in-progress**)

Thank you

Any questions?