Course Title:

Data-centric AI: transforming raw data into smart data

Teacher(s)
Vincenzo Pasquadibisceglie

Course description
The large amount of data currently collected in multiple domains (e.g., process management and monitoring, Earth observation, security) introduces additional challenges related to the resources needed for their analysis. Approaches for Big Data Analytics provide a solution for analysing such big amount of data by accounting for volume, variety, velocity challenges. However, the still suffer from the need for substantial computational resources. In this course, we will present effective methods to transform (big) raw data into smart data, representing summaries, conceptualisation, approximations, or even simplified and directly interpretable versions of the original ones. The course will consist of two parts:

(i) Design of methods for smart data extraction in data-centric AI: This part of the course will illustrate the state-of-the-art methods to extract smart data from raw data by resorting to data-centric AI approaches. The objective is to empower participants with the capacity to derive valuable insights and knowledge from raw data by handling volume and variety to achieve value in data. Smart data extraction encompasses procedures like sampling, simplifying concepts, and generating approximations.

(ii): Assessing the quality of data-centric AI: Understanding the quality of data-centric AI is pivotal to ensure the credibility of insights derived from smart data. In this course, participants will acquire the skills to assess the quality of smart data and their impact on AI-driven decision-making. Robust quality assessment mechanisms are crucial to guarantee the reliability and utility of smart data in the data-centric AI perspective. Through these assessments, participants will gain the proficiency to appraise the reliability of their smart data outputs and make informed decisions based on them.

Course period
November-December 2023

SSD
ING-INF/05

Course References (optional)

1. Predictive process mining meets computer vision. Pasquadibisceglie, V., Appice, A., Castellano, G., Malerba, D. Lecture Notes in Business Information Processing, 2020, 392 LNBIP, pp. 176–192;
2. T. Bai, J. Luo, J. Zhao, B. Wen, and Q. Wang, “Recent advances in adversarial training for adversarial robustness,” in 30th International Joint Conference on Artificial Intelligence, IJCAI 2021, 2021, pp. 4312–4321;

3. Gabriel M. Tavares, Rafael S. Oyamada, Sylvio Barbon, Paolo Ceravolo, Trace encoding in process mining: A survey and benchmarking, Engineering Applications of Artificial Intelligence, Volume 126, Part D, 2023.

Credits and Hours

3 credits, two of lectures (8 hours) and one of practice (15 hour), for a total of 31 hours.

Exam Modality

Two alternatives are available to the student to pass this exam:

1) Paper presentation. Students present the content of 2 papers suggested by the teacher. No groups are allowed.
2) Project. Students implement and experimentally validate an algorithm or its variation from a paper suggested by the teacher. Projects can be done in groups of 1-3 students, depending on the algorithm.

Teacher(s) CV

Attach or link a max 3 pages CV for each teacher proposing the course.

Teacher(s) Main Publications

1. DARWIN: An online deep learning approach to handle concept drifts in predictive process monitoring. Pasquadibisceglie, V.,Appice, A.,Castellano, G.,Malerba, D.Engineering Applications of Artificial Intelligence, 2023, 123, 106461;
2. STARDUST: A Novel Process Mining Approach to Discover Evolving Models From Trace Streams. Pasquadibisceglie, V., Appice, A., Castellano, G., Fiorentino, N., Malerba, D. ,IEEE Transactions on Services Computing, 2023, 16(4), pp. 2970–2984;
3. PROMISE: Coupling predictive process mining to process discovery. Pasquadibisceglie, V., Appice, A., Castellano, G., van der Aalst, W. Information Sciences, 2022, 606, pp. 250–271;
4. A Multi-View Deep Learning Approach for Predictive Business Process Monitoring. Pasquadibisceglie, V., Appice, A., Castellano, G., Malerba, D. IEEE Transactions on Services Computing, 2022, 15(4), pp. 2382–2395;
5. FOX: a neuro-Fuzzy model for process Outcome prediction and eXplanation. Pasquadibisceglie, V., Castellano, G., Appice, A., Malerba, D. Proceedings - 2021 3rd International Conference on Process Mining, ICPM 2021, 2021;
6. Orange: Outcome-oriented predictive process monitoring based on image encoding and CNNs Pasquadibisceglie. V., Appice, A., Castellano, G., Malerba, D., Modugno, G. IEEE Access, 2020, 8, pp. 184073–184086, 9216173;
7. Predictive process mining meets computer vision. Pasquadibisceglie, V., Appice, A., Castellano, G., Malerba, D. Lecture Notes in Business Information Processing, 2020, 392 LNBIP, pp. 176–192;
8. Using convolutional neural networks for predictive process analytics Pasquadibisceglie, V., Appice, A., Castellano, G., Malerba, D. Proceedings - 2019 International Conference on Process Mining, ICPM 2019, 2019, pp. 129–136, 8786066;

9. Contact-less real-time monitoring of cardiovascular risk using video imaging and fuzzy inference rules Casalino, G., Castellano, G., Pasquadibisceglie, V., Zaza, G. Information (Switzerland), 2019, 10(1), 9;
10. FISDeT: Fuzzy inference system development tool Castellano, G., Castiello, C., Pasquadibisceglie, V., Zaza, G. International Journal of Computational Intelligence Systems, 2017, 10(1), pp. 13–22.

Curriculum Vitae

PERSONAL INFORMATION

Family name, First name: Pasquadibisceglie Vincenzo
Researcher unique identifier(s) ORCID: 0000-0002-7273-3882
Date of birth: 10.03.1992
Nationality: Italian
URL for web site: https://persone.ict.uniba.it/rubrica/vincenzo.pasquadibisceglie

- **EDUCATION**

03.03.2022 PhD in Computer Sciences and Mathematics.
Department of Computer Science, University of Bari Aldo Moro, Italy.
PhD Supervisor: Annalisa Appice.

2015-2017 Master's degree in Computer Science.
Department of Computer Science, University of Bari Aldo Moro, Italy.

- **CURRENT POSITION(S)**

27.02.2023 Current Position: RTDa – Assistant professor (non-tenure track) - FAIR - Future AI Research

26.02.2026 (PE00000013), Spoke 6 - Symbiotic AI (CUP H97G22000210007), under the NRRP MUR program funded by the NextGenerationEU.
Department of Computer Science, University of Bari Aldo Moro, Italy.

- **PREVIOUS POSITION(S)**

02.02.2023 Position held: Occasional contract.

26.02.2023 Name of collaborators: Annalisa Appice.
Topic: Predictive Process Monitoring – KOMETA – Knowledge Community for Efficient Training through Virtual Technologies, Azione 1.4 "Promozione di nuovi mercati per l'innovazione" – Avviso pubblico INNOLABS- approvato con A.D. n. 13 del 08/02/2017, A.D. n. 37 del 28/03/2017 e A.D. n. 43 del 10/04/2017, tipologia "Knowledge Community").
Department/Institution: Department of Computer Science, University of Bari Aldo Moro, Italy.

06.12.2021 Position held: Research contract.

05.12.2022 Name of collaborators: Michelangelo Ceci, Donato Malerba.
Topic: Social media data acquisition and analysis - CounteR (European Project) - Privacy-First Situational Awareness Platform for Violent Terrorism and Crime Prediction, Counter Radicalisation and Citizen Protection, (Grant ID:101021607), European Union's Horizon 2020.

Department/Institution: CINI Consorzio Interuniversitario Nazionale per l'Informatica, Italy.

- **FELLOWSHIPS AND AWARDS**

07.12.2018 PhD XXXIV cycle in Computer Science and Mathematics
07.12.2021 Scholarship of University of Bari Aldo Moro - PON RI 2014-2020

2019 Vision and Imaging Technology Award: Call for Demos - 9th International Conference on Imaging for Crime Detection and Prevention – ICDP 2019 16-18 December, London UK

2022 BPM 2022 Best Dissertation Award, Doctoral Consortium, and Demonstration & Resources Track – PhD thesis shortlist: https://ceur-ws.org/Vol-3216/

- **ORGANISATION OF SCIENTIFIC MEETINGS**

13.11.2023 Local organizer of 3rd International Joint Conference on 15.11.2023 Learning & Reasoning (IJCRL), Bari – Italy

https://ijclr2023.di.uniba.it/~ijclr2023/organisers/index.html

- **INSTITUTIONAL RESPONSIBILITIES**

Co-supervisor of n. 10 bachelor thesis in Computer Science at Department of Computer Science, University of Bari Aldo Moro.
Topics: Predictive Process Monitoring, Computational Intelligence, Computer Vision.

- **REVIEWING ACTIVITIES**

Vincenzo Pasquadibisceglie has been a reviewer for several international journals, such as Expert System with Applications, IEEE Transactions on Services Computing, Journal of Grid Computing, etc. and is a member of the program committee of international conferences. Below are the most recent activities:

2023 Program Committee Member: ICDE 2024 (core rank A*): 40th IEEE International Conference on Data Engineering, 2024, Netherlands

2023 Program Committee Member: ECML PKDD 2023 (core rank A) - European Conference on Machine Learning and Principles and Practice of
Knowledge Discovery in Databases 2023, Turin, Italy

2022 Program Committee Member: ECML PKDD 2022 (core rank A) - European Conference on Machine Learning and Principles and Practice of
Knowledge Discovery in Databases 2022, Grenoble, France

- **MEMBERSHIPS OF SCIENTIFIC SOCIETIES**

Member, IEEE TFPM - Task Force On Process Mining

- **MAJOR COLLABORATIONS**

07.11.2020 Name of collaborators: Prof. Wil van der Aalst - RWTH Aachen University.

06.05.2021 Topics: Process Discovery, Predictive Process Monitoring
Department/Institution: RWTH Aachen University - PADS Laboratory
(https://www.pads.rwth-aachen.de/cms/~pnbx/PADS/lidx/1/)

2019-2021 Name of collaborators: Eng. Giuseppe Modugno – MTM Project srl.
Topics: Predictive Process Monitoring
Department/Institution/Company: MTM Project srl, Monopoli (BA), Italy
(https://www.mtmproject.com/)