Streaming Data Analysis
Administrative Items

POLITECNICO
MILANO 1863

Emanuele Della Valle

prof @ Politecnico di Milano

founder @ Quantia Consulting

founder & CRO @ motus ml

Me

Emanuele Della Valle, Ph.D.

prof @ Politecnico di Milano
founder @ Quantia Consulting
founder & CRO @ motus ml

+20 years of experience in **research** and **innovation** projects
Expert in Semantic Technologies, Stream Computing, and **Data Visualization**
Brander of **Stream Reasoning**: an approach to tame the velocity and variety dimension of Big Data simultaneously

Serial **startupper:**

- https://www.fluxedo.com/
- https://www.quantiaconsulting.com/
- https://motusml.com/

The other lecturers

Alessio Bernardo, Ph.D.

postdoc @ Politecnico di Milano
founder & CTO @ motus ml

Research in the **Streaming Machine Learning** field with evolving data streams, concept drifts, and class imbalance. Focus on applying Streaming Machine Learning techniques in constrained environments at the **network's edge**

Federico Giannini, Dr.

PhD student @ Politecnico di Milano

Researcher in the Streaming Machine Learning and **Continual Learning** fields. Focus on applying Deep Learning techniques to data streams to address concept drifts, temporal dependence, and catastrophic forgetting

Giacomo Ziffer, Dr.

PhD student @ Politecnico di Milano
founder & CEO @ motus ml

Researcher in the **Time-Evolving Analytics** field, focusing on applying Streaming Machine Learning techniques to (un)structured data streams with concept drifts and temporal dependence

The Course

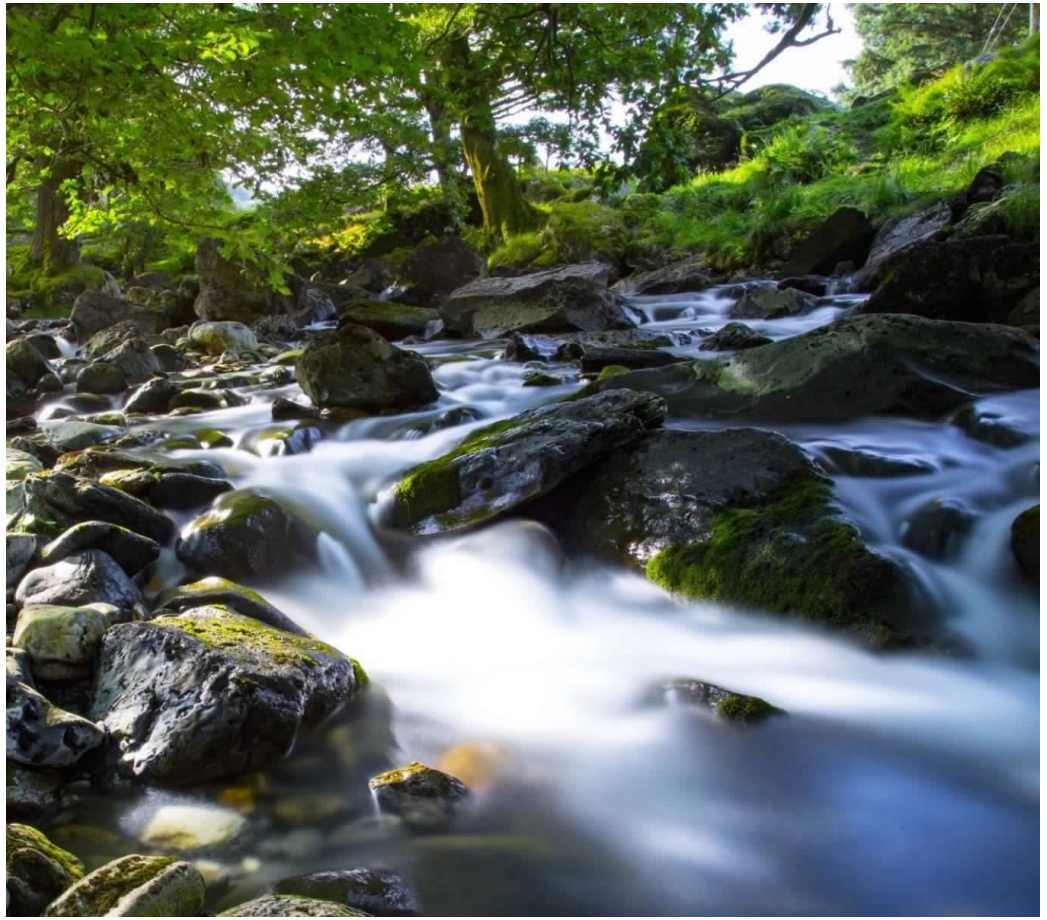
- Public course **Web** page with **official calendar of lecture**
  - http://emanueledellavalle.org/teaching/streaming-data-analytics-2023-24/
- Private **Webeep** page with official **recordings** and **announces**
  - https://webeep.polimi.it/course/view.php?id=11293
- Public **github** repo with all lectures' **slides** and **code**
  - https://github.com/emanueledellavalle/streaming-data-analytics

Lectures' Timetable

- **Official**
  - Wednesday 14:15 - 16:15 in classroom 26.11
  - Thursday 16:15 - 18:15 in classroom 3.0.3
- **Pragmatically**
  - Wednesday 14:30 - 16:00 in classroom 26.11
  - Thursday 16:30 - 18:00 in classroom 3.0.3
- **Important NOTES**
  - there is no clear cut between *theory* and *practice*
  - *bring your laptops* we will often code
  - I **record ALL lectures** but I do not stream them

Objetives

- The course provides the foundational **concepts**, **methods**, **languages**, and **systems for** ingesting, processing, and analyzing **data that flows** to enable real-time decisions.
- The course aims to **tame** the **velocity** dimensions of Big Data without forgetting the volume and variety dimensions.



Topics covered

Streaming Data Engineering

- From the foundations of streaming algorithms to real-world languages and systems
- Languages for Data Stream Management Systems (DSMS) and Complex Event Processing (CEP) illustrated via EPL
- Horizontally scalable DSMS illustrated via Apache Kafka, Apache Spark, and ksqlDB
- Vertically scalable CEP illustrated via Esper

Topics covered

Streaming Data Science

- From the foundations of Streaming Data Science to real-world Python libraries
- **Streaming Machine Learning** using River and MOA
- **Time Series Analytics** using statsmodels.tsa, and darts
- **Continual AI** using Avalanche

Evaluation

- The exam consist of three parts
  - a **mandatory written exam** (max 30 marks)
  - an *optional continuous evaluations* (max 1 mark)
  - an *optional practical project work* with oral presentation (max 3 marks)

- **Example**:
  - written exam 27 +
  - optional continuous evaluation 1 +
  - optional practical project work 3 =

30L

Evaluation

The "mandatory" written exam

- The written exam is composed of a mix of
  - **theoretical questions** regarding any course subjects and
  - **exercises** regarding the technical content and how to apply it in practice

- Students can get **up to 30/30** with the written exam

Evaluation

The "optional" continuous evaluations

- The optional continuous evaluations are in-presence quizzes proposed **during the lessons**
- I will use Microsoft Forms offered by Politecnico di Milano as part of your Office 365 subscription
- The quizzes are meant as a self-assessment; only participation matters

- Students can get up to **1 mark** with the continuous evaluations

Evaluation

The "optional" practical project

- The optional practical project requires using one or more of the technologies presented in the lectures to **solve a realistic** streaming data analytics **problem** based on data streams publicly available or provided by me and my assistants

- **Only students, who will get at least 27/30 in the written exam, can opt for it.**
  - **The mark obtained with the optional continuous evaluation does not count.**

- Students can get up to **3 marks** with the practical project

Streaming Data Analysis
Administrative Items

POLITECNICO
MILANO 1863

Emanuele Della Valle

prof @ Politecnico di Milano

founder @ Quantia Consulting

founder & CRO @ motus ml