

SmartData@PoliTO

https://smartdata.polito.it

Marco Mellia <u>marco.mellia@polito.it</u>

Big Data & Data Science & ML

Risk!

Design

(10100 (110100 (0191

Deposit



POLITECNICO DI TORINO

SmartData@PoliTO in one slide



Parent organization Politecnico di Torino

Sponsors
Both Private and public

Location **Torino**

Born in **2017**

Website smartdata.polito.it

Coordinator Marco Mellia

Address **Corso Duca degli Abruzzi 24 - 10129 Torino** Contacts **contact.smartdata@polito.it**









agilelab **Apprendistato** Supervisiona Impresa Assume √ddfor⁺ ScuDo ScuDo • Creano ponti Sviluppano Università **H-FARM** • Studiano IERRA INNOVATION





Hardware

SmartData @ PoliTO

- 1400+ CPU cores
- 19 TB RAM

Centro Nexa

- 8 PB storage
- 50 Gb/s network

Big data processing

HPC @ PoliTO

- 2900 CPU cores
- 27 TB RAM
- 26 V100 GPUs
- 100 Gb/s InfiniBand

$1+1[+1]^{\text{ting}}$

Cluster ~55TB RAM, ~6000 Cores, 10PB

Cloud @ PoliTO

- 1500 CPU cores
- 9 TB RAM
- 400 TB storage

Generic virtual machines

Uniti 44 Garage





POLITECNICO **Applications and domains DI TORINO** Predictive maintenance Models Virtual analysis Anomaly detection Making Cybersecurity Internet Energy Mobility in smart cities lization Product data Reconstruction Physics Signal Privacy 1Ce Security Data diffusion in complex Statistical Numerica Cybersecurity Management People & Market Culture utritional Strateo



JPC4AI

	ABOUT ~ EVENTS ~ SEMINARS ~	EDUCATION ~ RESEARCH ^
SmartData@PoliT0	OPEN POSITIONS V BIGDATA@POLI	PUBLICATIONS
		CURRENT PHD
HOME # PROJECTS		PROJECTS
Projects		OPEN DATASETS & LIBRARIES
/ EDIT		TOPICS

Current active projects

Data-driven quality assessment of espresso coffee production.



The project in collaboration with Lavazza aims at analyzing espresso coffee production data collected from professional coffee machines to improve the quality, lower the maintenance costs, and provide a better customer experience. Different quality variables are correlated to automatically extract human-readable patterns among the available data. Results include the data-driven definition of new and better quality values and the improvement of quality classification by means of time-series feature

engineering.

Contributors: Elena Baralis, Daniele Apiletti, Eliana Pastor

PIMCity – Building the Next Generation Personal Data Platforms



The Web economy has been revolutionized by unprecedented possibility of collecting massive amounts of user personal data, which lead the web to become the largest data market and created the biggest companies in our history. Unfortunately, this change has deep consequences



SmartData

for users, who, deprived of any negotiation power, are compelled to blindly provide their data for free access to services. Data collection is opaque,

fragmented and disharmonic, so that users have no control over their personal data, and, thus, on their privacy. Personal Information Management Systems (PIMS) aim to give users back control over their data, while creating transparency in the









Grigi

GN

Predictive Maintenance





Advanced Diagnostic/Predictive Maintenance





Not always things go right









10 years of data



Expectations vs Reality

Events

Fault



First law of big data: *Garbage in, garbage out*



NALYZE

000

D. Renga at al, (2020) <u>Data-driven exploratory models of an electric distribution network for fault prediction and diagnosis</u>, In: COMPUTING
 M. Nisi at al, (2019) <u>Transparently mining data from a medium-voltage distribution network: a prognostic-diagnostic analysis</u>, In: EDBT/ICDT Joint Conference





PRE-fault window

Virtual sensor & predictive maintenance











*e*EDISON

- Smart City
 - Harvesting of open data for the characterization of energy consumption of residential and commercial buildings





Mobility prediction

- Study of traffic patterns from car sharing systems
- Design of electric car sharing system





Number of units = 44

Model 1254 cluster 0 - Engine Percent Load vs fu



Service characterization, profiling, and shaping

 Analysis of IoT CAN bus data related to on-road and off-road vehicle usage







Internet traffic monitoring and security

 Analysis of internet traces looking for malware and other anomalies



- Digital transformation
 - Image analytics for near-duplicate detection



ML/A

processing

Actions to

manage

possible failure



- Failure detection, isolation and recovery for Space applications based on AI
 - Data from LEO satellites to detect&predict failures
 - Development of ML/AI algorithms with limited-resource hardware platforms





Lesson learned



- Data & ML work!
- You need a lot of sweet and experience to arrive to the results





• Data & ML work!

- You need a lot of sweet and experience to arrive to the results
- Data is dirty, terribly dirty & ML is complex, terribly complex. No Harry Potter magic wand

Lesson learned



Lesson learned

- Data & ML work!
- You need a lot of sweet and experience to arrive to the results
- Data is dirty, terribly dirty & ML is complex, terribly complex. No Harry Potter magic wand
- [cannot write it ©]



















https://smartdata.polito.it