DATACAMP

Thomas Moreau, MIND team @ Inria Saclay
thomas.moreau@inria.fr

Pedro L. C. Rodrigues, Statify team @ Inria Grenoble
pedro.rodrigues@inria.fr
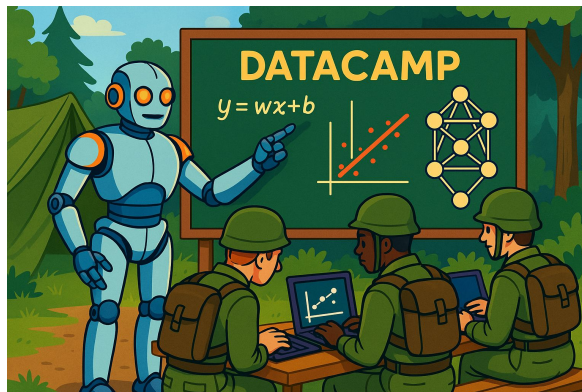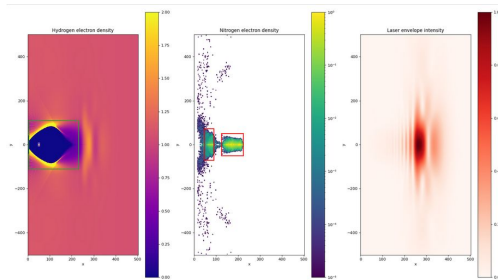
Objectives

Proper validation is the root for fast progress in data science

Put basic machine learning analysis knowledge to test by:

- **Solving practical data science problems** in scientific or industrial applications

- **Designing and evaluating data science workflows**

In the process, we also teach about:

- Practical tool boxes for various problems (missing data/unbalance/ …)
- Technical tools to improve your workflow (parallelism/debugging/ …)
- How to use proper tools to share code (git/github/formatting/…)
- Tools to test your code and build some documentation

Evaluation

- 50% of the grade comes from participation in the **data challenges** (Mostly based on performance on leaderboard) – **Individual** part

- 50% comes from a **team project** where students build a new data challenge (Teams of 6 students) – **Group** part

- Plus **three assignments** during the week, evaluated through **automated github evaluation**.
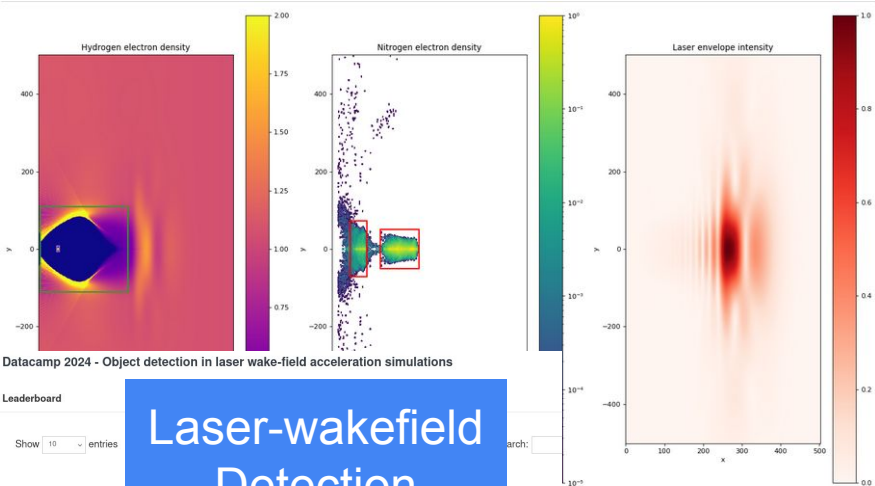
A practical class

Based on:

- Notebooks
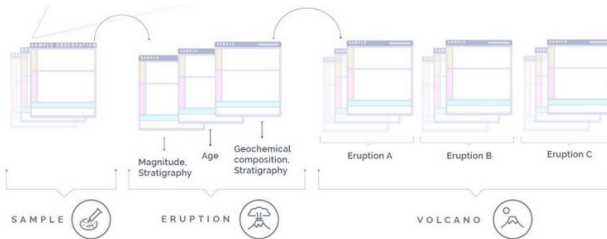- Practical sessions
- Challenges

Challenge examples

Leaderboard

Show 10 entries

Search:

rank	move	team	submission	mean_absolute_error	train time [s]	validation time [s]	test time [s]
1	-	bryanbradfo	xgggggggggggggggggg	14.26559	563.942311	748.780373	255.972701
2	+1	Bapt	test_add_features_2	14.32436	812.484430	1217.861139	425.912225
3	+6	emmadchrr	sub_060325_EDC	14.33353	10282.262520	122.788028	42.212423
				14.37971	1679.645033	14.686541	4.722813

Mean Arterial Pressure estimation

Hydrogen electron density

Nitrogen electron density

Laser envelope intensity

Leaderboard

Laser-wakefield Detection

Show 10 entries

Search:

rank	move				test time [s]
1	-	AyanAhmed	submission17	0.948	33604.639820	844.152903	209.679316
2	+1	akumar	sub#3	0.939	11796.633674	413.969064	102.938337
3	-1	rayyan_ahmed	submission_18	0.938	33003.745918	827.775765	205.617334
4	+6	nourhezbri	sub_2	0.935	36738.218505	398.129796	98.875122
5		marssoo	model4-7	0.934	79636.256295	414.257084	103.109875

SAMPLE — ERUPTION — VOLCANO

Magnitude, Stratigraphy

Age

Geochemical composition, Stratigraphy

Eruption A

Eruption B

Eruption C

Linking eruption to volcano in tephra

...erboard

ow 10 entries

rank	move	team	submission	bal_acc	train time [s]	validation time [s]	test time [s]	submitted at (UTC)
	-	Idriss-M	Starting_kit0	0.805	1228.573692	5.305873	0.998519	2024-02-17 11:33:52
2	+34	Aymane-Rahmoune	train_test9	0.799	1900.692650	4.948802	0.771890	2024-03-03 21:24:23
3	+24	Fede-Testa	smotek_med_nobal	0.791	699.307611	2.218053	0.354542	2024-03-10 17:18:14