

RAMPs: CHALLENGES WITH CODE SUBMISSION

BALÁZS KÉGL

LAL / CNRS

ALEXANDRE GRAMFORT

LTCI / Telecom ParisTech

ISABELLE GUYON

LRI / UPSud

AKIN KAZAKCI

Ecole des Mines

CAMILLE MARINI

LTCI / CNRS

MEHDI CHERTI

LAL / CNRS

WHY CODE SUBMISSION?

- To make it **collaborative**
 - but: to make it useable for student evaluation, we **need a closed phase**
- To tackle **complex workflows**
 - time series, DAG-like workflows, data augmentation
 - but: to make it easy to use, we need to **make contribution simple**
- Quite complex **engineering challenge**

IN THE CLASSROOM

- Introduction

- presenting the **use case**, the scientific or business problem
- presenting the **starting kit**, typically an iPython notebook with **exploratory analysis** and the **analytics pipeline**, example **submissions**, the evaluation **metrics**

- Closed phase

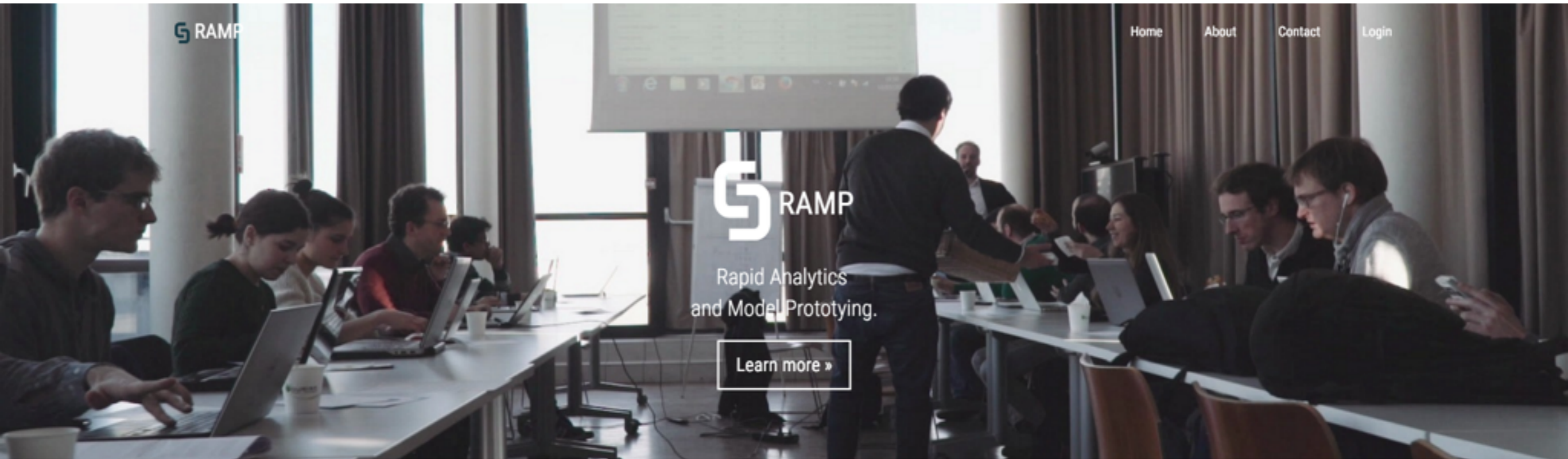
- typically **1-2 weeks when students work alone**, **submit once/twice a day**, we coach/consult them in the classroom
- we **mark** them by their (absolute) **performance** at the end of the closed phase plus a **report** in the form of an iPython notebook

- Open phase

- typically a **single day “hackaton”**, all **code is reusable**
- students **present their ideas** during the day

RAPID ANALYTICS AND MODEL PROTOTYPING (RAMP)

<http://www.ramp.studio>



Collaborative prototyping

During the RAMP, the participants submit predictive solutions (code). The models are trained on our back-end. The scores are displayed on a leaderboard. All participants have access to all code, and they are encouraged to look at and to reuse each other's solutions. This accelerates the development process since good ideas spread fast.



Training

A great tool to learn data science! RAMPs are used in the MS Big Data at Telecom ParisTech, in three UPSaclay M2 programs (Data Science, AIC, Data and Knowledge), in a course on Machine Learning for Finance and Economics at Université Panthéon-Assas, in a graduate course in the Data analysis and decision program at Ecole Centrale de Lille.



Networking

Each RAMP attracts about 30-50 participants, coming from different backgrounds and carrier stages, who usually meet for the first time. They develop a working relationship in a relaxed environment, and sometimes keep working together after the event.

<http://www.ramp.studio>

- Already **open for limited trials**, ping me if you would like to try it
 - El Nino, Arctic Ice: **spatio-temporal forecasting**
 - Drug spectra: **functional data**, **DAG-like workflow**
 - Air passengers: **data augmentation**
 - HEP anomaly: **noisy classification**
 - Epidemium: **multi-target regressions**, tons of **missing data**

QUESTIONS FOR THE DISCUSSION

- How do you use **challenges in the classroom**?
- How to **evaluate students**?
- **Overfitting** in challenges?
- What **features** and **extensions** would be interesting to add?
- Competition/collaboration/coopetition?
- **Reproducible** research?
- **Gamification**?
- Engaging **women** in AI/ML?
- **Machine/human** collaboration?