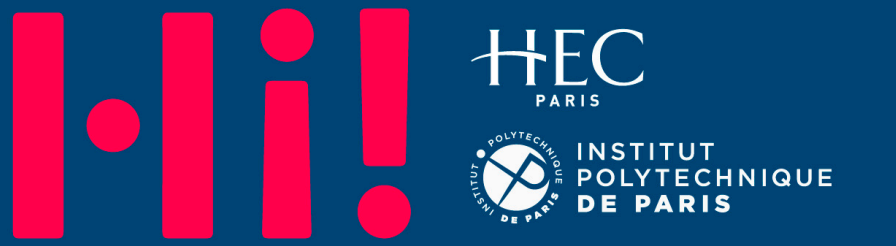


PARIS ARTIFICIAL INTELLIGENCE FOR SOCIETY

Hi!ckathon 2023

Student Guide

January 2023

AGENDA

1. Hi!ckathon description
2. Understanding the subject
3. Work methods
4. Deliverables and deadlines
5. Coaching
6. Timeline of the Hi!ckathon

What is the Hi!ckathon ?

The event

Competitive Data / AI event of Hi! Paris, the interdisciplinary research and teaching Center for Data Analytics and Artificial Intelligence

The theme

AI & Sustainability

Beyond producing a Data / AI model, the competition will ask to realistically project the solutions in a market context

When?

From Friday **January 13th, 6:30pm** to Monday **January 16th, 7pm**

The Partners

Corporate donors

L'ORÉAL · Capgemini · TotalEnergies · KERING · Rexel · VINCI · Schneider Electric

Their role

- Enrich the event with their **business expertise**
- Share **concrete issues** they have faced

Their contribution

- Participation in **webinars** to **discuss real use cases** on AI and Sustainability (broadcasted on Zoom)
- **Introduction of sustainability's challenges**

Closing ceremony

- Present the **seven thematic prizes** which will reward the winning teams of the competition

What will you have to produce during the event?

- Build a **Data / AI model** addressing the problem 
- Build a **scientific procedure document** to justify your approach 
- To put the AI solution in context, your team will have to produce a **video pitch** that illustrates the project in terms of **business and/or societal impacts** 

*To help you achieve these objectives, **coaching** involving business **experts** and **data scientists** will be set up (see page 17)*

AGENDA

1. Hi!ckathon description
2. Understanding the subject
3. Work methods
4. Deliverables and deadlines
5. Coaching
6. Timeline of the Hi!ckathon

« Modelling the energy consumption of buildings to support energy transition »

Subject

Context

Despite the importance of real estate in GHG emissions and energy consumption, investments in energy transition for the housing sector remain insufficient.

Objective

Giving real estate owners advice and more visibility on the benefits from insulation works would be a strong enabler for energy transition investments.

Data

A dataset of 1.5 million European buildings, with their technical features and yearly energy consumptions.

Purposes

Business perspective

In order to reduce their carbon footprint, real estate owners need to:

(i) Quickly measure their property's impact

(ii) Simulate the energy and cost savings in various scenarios to determine the best strategy

Pedagogical point of view

- Machine learning: Feature Engineering, feature selection, handling large datasets
- Best practices: PEP8 code, code structure (organized in methods & classes)
- AI project management: developing code in community using CI/CD tools efficiently
- Business: technology at the service of business, the art of pitching ideas on deploying the algorithmic solution

Expected deliverables

- Data model
- Scientific procedure document
    - Rationale for the Data Science approach
    - Best practice, frugality of the code
- Business pitch

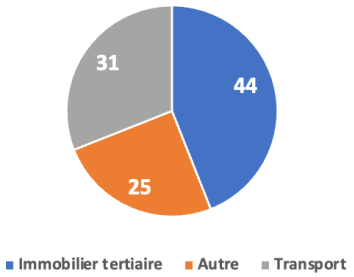
Hackathon subject: Modelling the energy consumption of buildings to support energy transition

Context

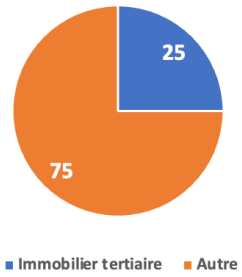
Real estate is a key contributor to GHG emissions…

Share of real estate…

In France's total energy consumption

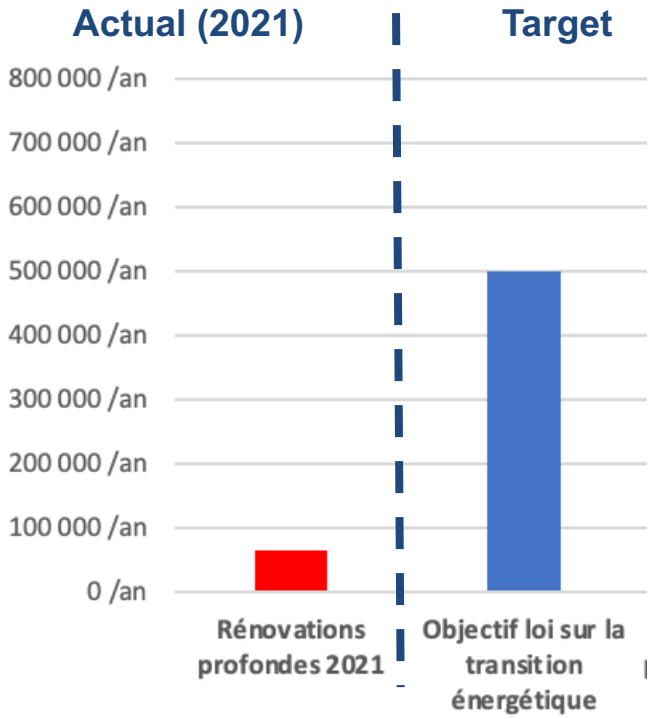


In France's GHG emissions



… yet investments in thermal insulation lag far behind objectives

Insulation works in France p.a.



Objective

ML task

Predict the energy consumption of European buildings in order to estimate energy and cost savings, and automate energy assessments.

Business goal

Prepare a business pitch for an app that:

(i) Leverages your ML model

(ii) Addresses the key obstacles to refurbishment

Data

A dataset of 1.5 million European buildings, with their technical features and yearly energy consumptions

Agenda

1. Hi!ckathon description
2. Understanding the subject
3. Work methods
4. Deliverables and deadlines
5. Coaching
6. Timeline of the Hi!ckathon

Work efficiently

Build a clear, realistic and precise timeline

Plan your deliverables, set up provisional deadlines, and readjust

Take time to brainstorm and define roles in your team

Communicate frequently on your progresses (within your team and with coaches)

Define responsibilities… and adapt if necessary

- Represent the team and act as a final decision point when there is a divergence
- Keep the team to schedule
- Hardware and software engineering members
- Responsible for user experience design
- Define business modelling, targets, market
- Deep dive into the topic accurately
- Responsible for the delivery of video & presentation elements

Tools you use

- → The challenge topic and datasets are accessible via the HEC data platform Hfactory
- → GITLab
- → Development environments (JupyterLabs, Jupyter notebook classic, Theia)

Discord

- → Used to benefit from coaching rooms
- → Allows team members to communicate with each other via dedicated rooms
- → Share the latest news and announcements on the Hi!ckathon

AGENDA

1. Hi!ckathon description
2. Understanding the subject
3. Work methods
4. Deliverables and deadlines
5. Coaching
6. Timeline of the Hi!ckathon

Deliverable 1. Data Science model & predictions

Your task : building a frugal and performant model

Which scoring metric is used ?

$$explained\_variance(y, \hat{y}) = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}}$$

How are code quality & frugality evaluated?

Code quality:

- Executability : At the end of the challenge, you will be asked to execute your code (inference + preprocessing on the test set)
- Code structure : good use of methods / classes / modules to structure code, and docstrings to comment on your code
- Good use of Gitlab : we recommend not to include large data files in your repository, and to use branches and merge requests adequately

Frugality:

- Inference CPU time : measures if your model is heavy (thus not green) when used to make real-time decisions

Deliverable 2. Scientific procedure document

A clear & concise document

Mandatory sections

1. Overview

- ✓ Project Background and Description
- ✓ Project Scope
- ✓ Presentation of the group
- ✓ Task Management

2. Project Management

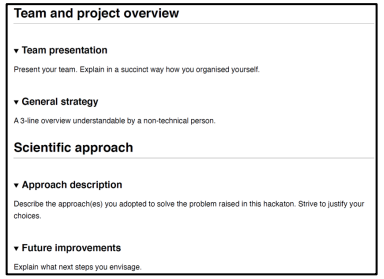
- ✓ Data Understanding
- ✓ Data Pre-processing
- ✓ Modeling Development
- ✓ Deployment Strategy

3. Carbon Footprint Limitation

4. Conclusion

Requirements / Target

- ✓ A **reading time** not exceeding **2 minutes**
- ✓ Submitted directly **within your GitLab : project** as part of the README file
- ✓ You will find the submission template in the folder Welcoming_final_challenge on the HFactory.



Click and check our template

Not filling a mandatory section is eliminatory for the scientific procedure award !

Deliverable 3. Video presentation

Be creative, be bold, be impactful

Requirements

Prepare a business pitch for an app that:
- Leverages your ML model
- Addresses the key obstacles to energy transition in the housing sector

Duration :
2 min

Language:
English

Tools you may use… or any other

Video creation

Animaker

Adobe Spark Video

Animiz

PLAY PLAY

Video editing

LIGHTWORKS

HITFILM

iMovie

Evaluation criteria

- **Communicate clearly** and **distinctively** to pitch the solution convincingly
- Structure your argument: **tell a story**, capture the audience and present your strategy
- Project themselves into a **business and/or societal framework**
- Imagine a **feasible project** with a **strong business opportunity**
- **Address clearly the subject** and its challenges

Key features for a compelling video presentation

The video presentation should …

- Provide us with a brief, clear wrap-up of your ideas
- Show the business and/or societal impact of your project
- Be no longer than 2 minutes
- Have a strategy to leverage the AI solution

What should the video contain ?

- A short presentation of your product and how it responds to the stake of sustainability
- A demonstration of the market potential of the idea. Here is a suggestion on how to do so (feel free to pick on some items and/or choose another approach):
  - Your team (who you are, how you are structured, your expertise & capabilities)
  - Business Idea (what type of product/service you plan to offer, why, to whom, what for and how)
  - Market (market trends & expectations, clients & persona, market size, market growth rate)
  - Go-to-market & sales (competition, pricing model, profitability as in ROI for example, distribution method, sales model, promotion strategy, financial forecasts, etc.)
  - Operations (IT and management information systems, premises, etc.)
- A discussion on the feasibility, effectiveness, efficiency and/or viability of the project *(optional)*

Deadlines and submission process

When should you submit your project?

- Submit all your files by **Sunday January 15th, 4pm.**
- **Between 4 and 5 pm, DS coaches will ask you to re-run your code (preprocessing of test set + inference), to assess code executability & quality**
- **Reach out if you have any problem !**

How should you submit your files?

The Data science model and scientific procedure document will be submitted through your GitLab project

- **Only the notebook in your JupyterHub instance within the HiParis DataFactory will be evaluated.**

The predictions of your model will be submitted and graded via the HFactory

Your video exploring business and/or societal opportunities will be uploaded on a different platform

- A Microsoft Forms link **will be sent to you during the event** to submit your video

Important note on plagiarism

- Teams are expected to only submit predictions obtained using their own machine learning models on the provided test set
- As a result, **the following situations may lead to elimination** :
  - Use of another team's code or prediction files
  - Discrepancies between the prediction file submitted via the HFactory and the actual output of a team's model during the correction phase (between 4 and 5 pm on Sunday)
- Please also note that, in case of identical submissions, **all involved teams will be excluded from the competition.**

AGENDA

1. Hi!ckathon description
2. Understanding the subject
3. Work methods
4. Deliverables and deadlines
5. Coaching
6. Timeline of the Hi!ckathon

Coaching

Throughout the Hackathon, coaches will be available to help you on your journey and create a solution for the challenge of your choice…

Who are the coaches?

Data Scientists 

- **Guide you** on your choices & propose leads
- Help **understanding datasets**
- Help in terms of methodology for constructing your **AI / Data solution & your report**

Consultants 

- Team & project **organisation**
- **Methodology** for the pitch production
- **Continuous support** for any enquiry / difficulty / specifying-rule needs

How can I request meeting a coach?

Data scientists & Consultants

- Available at **any time** (9am – 7pm) during the Hackathon

Experts

- Available on **specific slots**
- Need to be chosen **if relevant** to your subject
- **Visio** by specific team or grouped **Q&A** (depending on the amount of request)

What the coaches **cannot** do…

- Give ready-made solutions
- Intervene on the code

Your technical support team

Quentin de Lovinfosse
Managing consultant
Capgemini

Jean-Nicolas Vizy
Data Scientist
Capgemini

Solenna Ishac
Consultante
Capgemini

Baptiste Calot
Data Scientist
Capgemini

Here to:

- ✓ Answer your concerns about datasets
- ✓ Guide you to frame your solution
- ✓ Assist you to build your solution and the deliverables

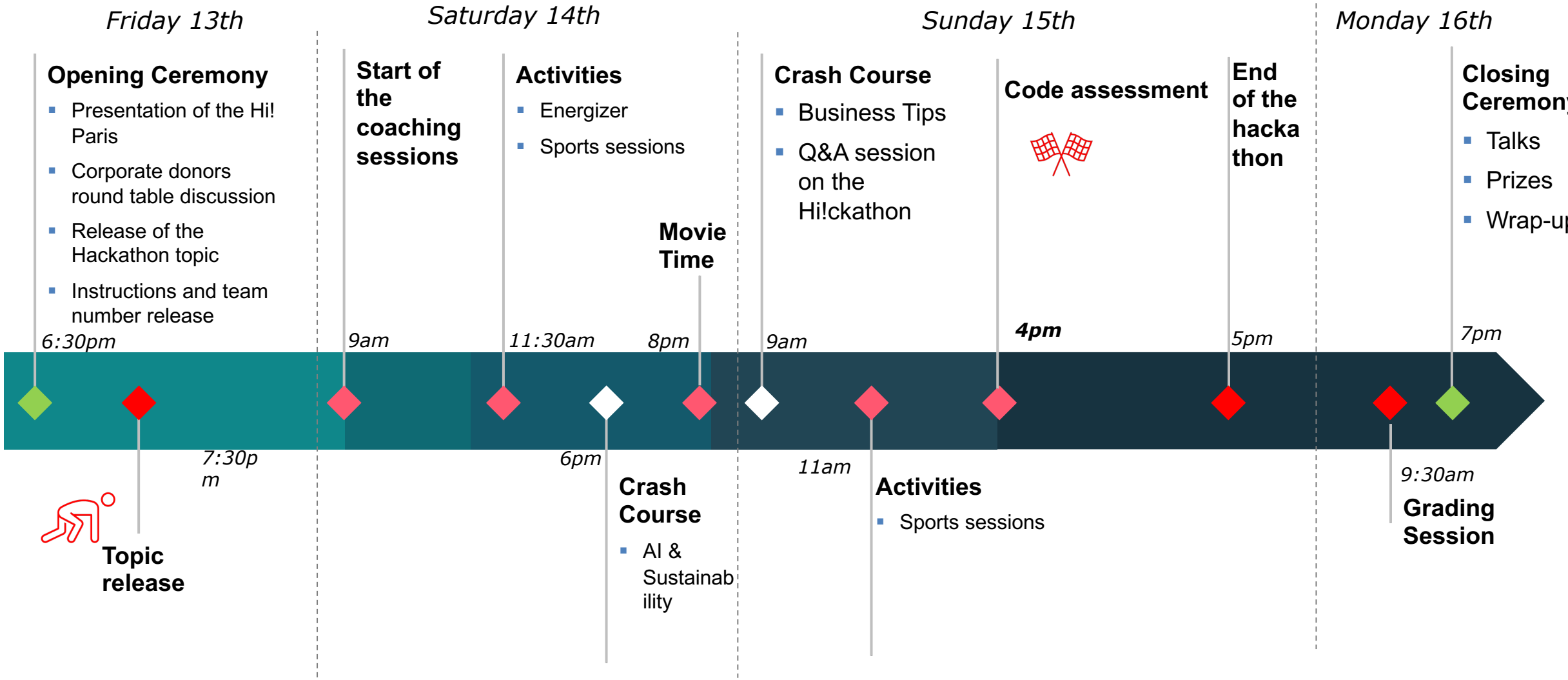
Available:

On all dedicated Discord channels

AGENDA

1. Hi!ckathon description
2. Understanding the subject
3. Work methods
4. Deliverables and deadlines
5. Coaching
6. Timeline of the Hi!ckathon

Timeline of the Hi!ckathon



All activities will be held only on site