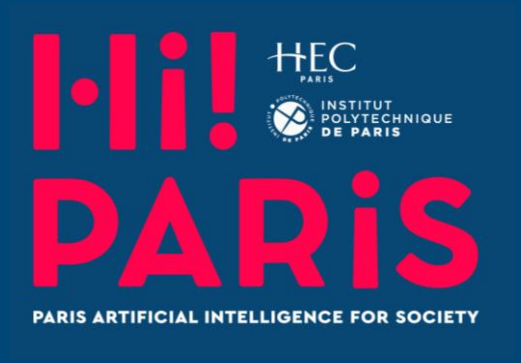


Hi!ckathon 2024

Student Guide

HEC PARIS

INSTITUT POLYTECHNIQUE DE PARIS

Inria

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 & Supply Chain

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

When?

From Friday **December 1st, 6:30pm** to Monday **December 4th, 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 Round Tables to **discuss real use cases** on AI and Sustainability (broadcasted on Zoom)
- **Introduction of Supply Chain 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*

AGENDA

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

SUPPLY CHAIN FOR SALES PREDICTIONS: THE CASE OF SCHNEIDER ELECTRIC

Dataset

A data set of 2.2 million lines, each containing 4 months of sales of a defined product, as well as extensive information including country of production, sales, target customer, macro-economic and environmental data.

Story Telling

You're a start-up specializing in the supply chain, you've received an anonymized dataset, along with various indicators (4 sub-datasets).
You must cross-reference these data (you're free to add to them) to predict the fourth month of each "third of a year". You'll also need to explain how these predictions were made and explain any variations. In addition, you must avoid overproduction and propose measures to reduce the overall carbon footprint.

Objectives and Deliverables

ML task
Predict sales volume for month 4

Scientific paper
Using Data Science, you must propose a solution to optimize sales prediction while taking into account the full CO2 emissions of the value chain.

Business goal
Design an app that:
1. Leverages your ML model
2. Present your result in sales prediction
3. Provides insights to improve sustainability

What we want

Sustainability Imperative
Holistic Approach
Innovation in Supply Chain Management
Collaboration for Impact
State-of-the-art Machine Learning

Agenda

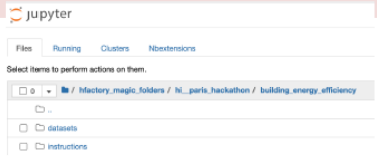
1. Hi!ckathon description
2. Understanding the subject
3. Work methods
4. Deliverables and deadlines
5. Coaching
6. Timeline of the Hi!ckathon

AVAILABLE RESOURCES & EXPECTED DELIVRABLES

WHAT DO I HAVE?

- Welcoming Notebook
- Student Guide
- Scientific procedure document template
- Datasets

README file



WHAT DO I SUBMIT?

- Data Model
- Scientific procedure document
- Video of the business pitch

A CSV file

Rationale for the Data Science approach

A 2 minutes video in English that shows the business and/or societal impact of the AI project

/ ··· / group_00002_shared_workspace / submissions /

Intellectual Property of the participants' code: We adopt an **open-source** distribution logic, according to which your AI solutions can be shared in a public GIT of Hi! PARIS

TOOLS TO USE

Access to the challenge topic and datasets

Factory

Communicate

zoom

Code your model

python

jupyter

GitLab

Edit a video

Animaker

Animiz

PLAY PLAY

Adobe Spark Video

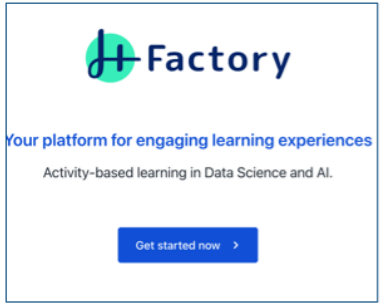
iMovie

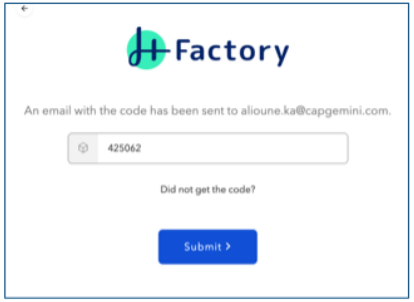
LIGHTWORKS

HITFILM

STEP-BY-STEP GUIDE ON THE HFACTORY PLATFORM

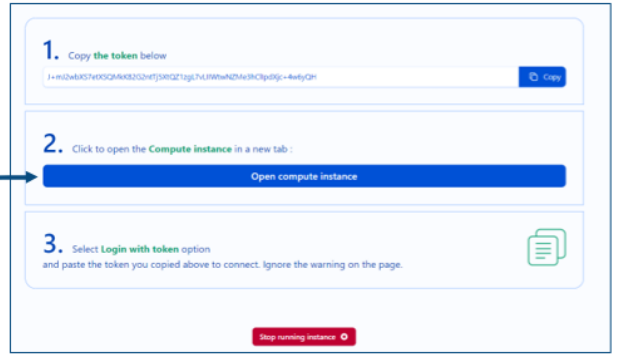
1. **Register to the Hi!ckathon #4 activity on HFactory** (you should have received an invitation by email)





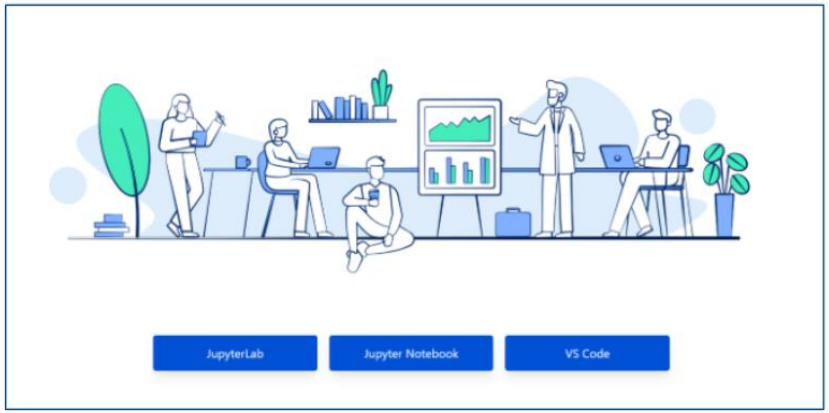
2. **Click on the platform's "Data Science" page to access the Python environment**

3. **Start a CPU/GPU computing instance** (note that GPU resources are limited) Then, copy the token then click on "**Open compute instance**" to login with it.

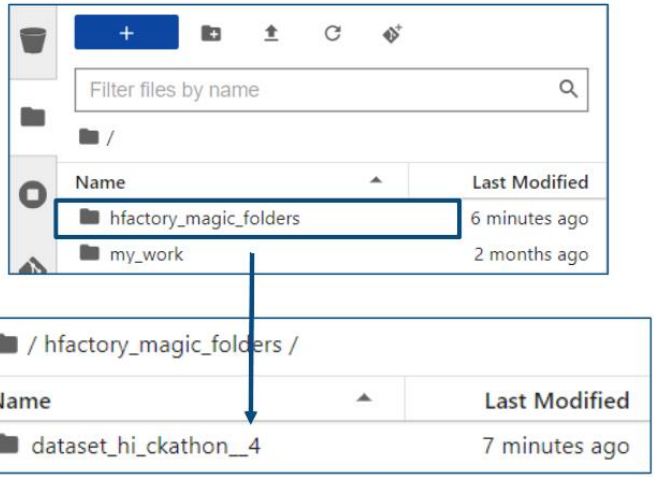


STEP-BY-STEP GUIDE ON THE HFACTORY PLATFORM

4 **Now, you can select an IDE to use with the computing environment** (JupyterLab, JupyterNotebook, VS Code)



5 Once you've selected an IDE, you can **open the computing environment**. In this environment, you can access the dataset and work in your own or group environment.



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

Hfactory

- → 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 ?

A **normalized version of the Root Mean Square Error (RMSE)** will be used. Its values are between 0 and 1. (0 and 100 on Hfactory)

Here is a formula explaining how the original RMSE is computed

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are predicted values

$y_1, y_2, \ldots, y_n$ are observed values

n is the number of observations

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

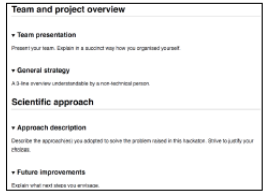
1. Overview
- ✓ Project Background and Description
- ✓ Project Scope
- ✓ Presentation of the group
- ✓ Task Management

2. Project Management
- ✓ Data Understanding
- ✓ Data Pre-processing
- ✓ Modeling Development
- ✓ Deployment Strategy

3. Sales Forecasting and Sustainability
4. Conclusion

Requirements / Target

- ✓ A **reading time** not exceeding **2 minutes**
- ✓ Submitted directly **within your GitLab : project** as part of the README file


Team and project overview

▾ Team presentation

Present your team. Explain in a succinct way how you organised yourself.

▾ General strategy

A 3-line overview understandable by a non-technical person.

Scientific approach

▾ Approach description

Describe the approach(es) you adopted to solve the problem raised in this trackator. Strive to justify your choices.

▾ Future improvements

Explain what next steps you envisage.

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 **December 3rd, 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?

Hi! PARIS Research Engineers and Data Scientist Consultants

- **Guide you** on your choices & propose leads
- Help **understanding the datasets**
- Help in terms of methodology for constructing your **AI / Data Science & your report**
- Team & project **organization**
- **Methodology** for the pitch production

Technical/Business experts

Benefit from technical/business advice from experts (Hi! PARIS Corporate Donors members) for any **Data Science questions** you may have or to gain insights on how to construct your business pitch.

How can I request meeting a coach?

Hi! PARIS Engineers and Consultants

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

Experts

- Available on **specific time slots** (Saturday and Sunday morning)
- Each team will be assigned a **Business/Technical coaching** session with an expert
- Not every team member needs to attend a session

What the coaches **cannot** do…

- Give ready-made solutions
- Intervene on the code

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


HEC
PARIS
INSTITUT
POLYTECHNIQUE
DE PARIS
PARIS
PARIS ARTIFICIAL INTELLIGENCE FOR SOCIETY