Papermill: Automatisation et paramétrage de notebooks jupyter





Notebooks jupyter

C'est super, mais limité dans certains cas d'usage:

- Exécuter un notebook avec des paramètres différents (hyperparamètres d'un modèle, valeurs input fluctuantes)
- Exécuter plusieurs notebooks automatiquement

Notebooks jupyter

C'est super, mais limité dans certains cas d'usage:

- Exécuter un notebook avec des paramètres différents (hyperparamètres d'un modèle, valeurs input fluctuantes)
- Exécuter plusieurs notebooks automatiquement

napermill 🏟

- Facile à utiliser
- Injecter des paramètres dans un notebook Jupyter
- Exécuter un/des notebook/s:
 - Python API
 - Ligne de Commande

Notebooks jupyter

C'est super, mais limité dans certains cas d'usage:

- Exécuter un notebook avec des paramètres différents (hyperparamètres d'un modèle, valeurs input fluctuantes)
- Exécuter plusieurs notebooks automatiquement

🃦 papermill

- Facile à utiliser
- Injecter des paramètres dans un notebook Jupyter
- Exécuter un/des notebook/s:
 - Python API
 - Ligne de Commande

Welcome to papermill

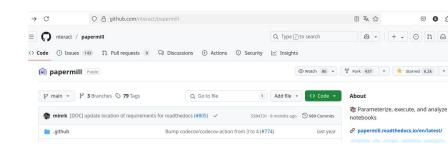


This opens up new opportunities for how notebooks can be used. For example:

- Perhaps you have a financial report that you wish to run with different values on the first or last day of a month or at the beginning or end of the year, using parameters makes this task easier.
- Do you want to run a notebook and depending on its results, choose a particular notebook to run
 next? You can now programmatically execute a workflow without having to copy and paste from
 notebook to notebook manually.

Python Version Support

This library currently supports python 3.7+ versions. As minor python versions are officially sunset by the python org papermill will similarly drop support in the future.



Papermill utilisation

Une cellule tag "parameters"

```
# input date

year = 2023
month = 10
day = 15

x parameters + Tag Python
```

Exécution du notebook en ligne de commande avec modification d'un paramètre:

papermill -p nom_parametre valeur_parametre nom_input_notebook.ipynb nom_output_notebook.ipynb

Papermill utilisation

Une cellule tag "parameters"

```
# input date

year = 2023
month = 10
day = 15

x parameters + Tag Python
```



Répo github

Exécution du notebook en ligne de commande avec modification d'un paramètre:

papermill -p nom_parametre valeur_parametre nom_input_notebook.ipynb nom_output_notebook.ipynb

Avec Papermill

Exécuter un/des notebooks automatiquement (ligne de commande, script python ou bash)

Workflow avec paramétrage possible selon l'output du 1er notebook

Exécuter tous les notebooks d'un dossier

Possibilité d'exporter le notebook exécuté en html via nbconvert

Ex de cas d'usage : Utiliser un notebook template pour générer des rapports html, tester différentes valeurs d'input ou jeux de données, modifier les hyper paramètres d'un modèle...

Liens

Documentation Papermill: https://papermill.readthedocs.io/en/latest/

Répo Github: https://github.com/nteract/papermill

Articles:

- → Automating Jupyter Notebooks with Papermill and Crontab https://medium.com/data-tyro/automating-jupyter-notebooks-with-papermill-and-crontab-39ea73218c96
- → Parameterizing and automating Jupyter notebooks with papermill: https://www.wrighters.io/parameters-jupyter-notebooks-with-papermill/

Alternatives

Ploomber: https://docs.ploomber.io/en/latest/index.html

Marimo: https://marimo.io/

Merci!