

Mapping learning



A.THOMAS
S.CORGNE – T.CORPETTI

Mapping learning, pourquoi ?

UMR LETG Rennes - COSTEL

- Climatologie
 - Cartographie (Occupation du sol)
 - Télédétection
 - Informatique
 - Géomorphologie
 - Ecologie... (j'en oublie ?)
- Géographes et informaticiens travaillent ensemble
 - ... et ça marche plutôt bien !
 - Quelques difficultés pour échanger leurs données, techniques et résultats

Et pourquoi j'ai pensé à ce projet ?

- Fonction d'accompagnement, de support pour la(les) recherche(s)
→ solution pour aider le plus grand nombre !



La cartographie

Avant-hier (la carte de Cassini)

- Relevés topographiques
- Début : fin XVIIIe s. - Fin : XIXe s.
- Auteur : Cassini (4 générations)
- 154 feuilles + 26 feuilles partielles



Aujourd'hui

- Une image (numérique)
- Durée : ordre de la journée (temps de traitement)
- Une carte numérique (raster ou vecteur)



Machine Learning, pourquoi ?

Pour un informaticien, 1 image = 1 matrice

- logiciels/codes qui travaillent sur des matrices
- Ne sont pas gérés :
 - La référence spatiale (où est-on ?)
 - Le système de projection (comment l'image est mise à plat sur le globe terrestre ?)

Le géographe reçoit ainsi un résultat

- Dans un format qu'il ne peut pas utiliser directement
- Auquel il doit réaffecter la dimension géographique



Mapping learning, pourquoi ?

Pour les informaticiens

- Accès simplifié à données de natures variées : tables, raster, vecteur (géographiques ou pas)
- Mise à disposition de leurs algorithmes
 - pour tests, avant proposer une intégration dans scikit-learn ?

Dans *Mapping Learning*

19 algorithmes de classification supervisée (merci scikit-learn !)

Validation croisée

courbe ROC

réductions de dimensions

Mapping learning, pourquoi ?

Pour les géographes et non-experts en Machine Learning

- Assiste dans chaîne de traitements
 - données (images de télédétection) → résultat (carte)
- Nombreux algorithmes de Machine Learning
 - Ex : 19 algorithmes de classification supervisée (merci scikit-learn !)
 - Pas tous implémentés dans logiciels de télédétection
 - Recherche des paramètres optimaux
- Rapport standardisé, résumant toutes les étapes et les résultats
 - Donne conseils d'ordre statistique, et empiriques

Mapping learning, pourquoi ?

Pour les informaticiens

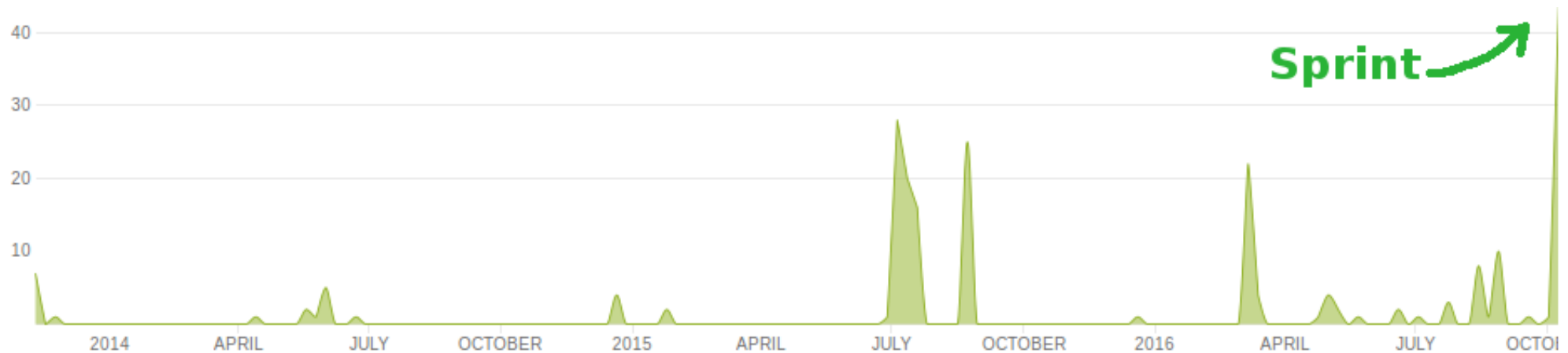
- Accès simplifié à données de natures variées : tables, raster, vecteur (géographiques ou pas)
- Mise à disposition de leurs algorithmes
 - pour tests, avant proposer une intégration dans scikit-learn ?

En plus dans *Mapping Learning* :

- Validation croisée
- Courbe ROC
- Réductions de dimensions
- Métriques de distance différentes (certains temporelles)

Mapping learning origin

- 2012 : 1 besoin = comparer n algorithmes de classification (à répéter x fois)
- Aides ponctuelles pour besoins similaires
- ↑ ambitions : capacité à utiliser différentes natures de données
- Création d'un dépôt (privé) fin 2013
- ↑↑ ambitions : + classification non supervisée et régression
- ↑↑↑ ambitions : présenter *Mapping Learning* au Pycon-fr !



Mapping Learning : état actuel

```
Totals grouped by language (dominant language first):
python:          3158 (100.00%)

Total Physical Source Lines of Code (SLOC)           = 3,158
Development Effort Estimate, Person-Years (Person-Months) = 0.67 (8.03)
  (Basic COCOMO model, Person-Months = 2.4 * (KSLOC**1.05))
Schedule Estimate, Years (Months)                   = 0.46 (5.52)
  (Basic COCOMO model, Months = 2.5 * (person-months**0.38))
Estimated Average Number of Developers (Effort/Schedule) = 1.46
Total Estimated Cost to Develop                      = $ 90,370
  (average salary = $56,286/year, overhead = 2.40).
```

generated using David A. Wheeler's 'SLOCCount'.

**Un bébé de 3k lignes et qui peut encore grandir
(mais pas trop vite)**

Mapping Learning : développement

En Python (!)

Principales librairies

- scikit-learn (+ mipy) : machine Learning
- osgeo (Gdal/Ogr) : fichiers géographiques
- Numpy & pandas : manipulation de données

Développement

- A.Thomas : développeur principal (vous êtes les bienvenus!)
- Approche orientée-objet
- Versionnement : Git (dépôt sur Bitbucket)
- Bonnes pratiques : PEP8, tests unitaires (71)



Mapping Learning est une application

Libre

Licence IGPL

Basée sur des logiciels et des bibliothèques libres



Utilisable par tous

Compatible Linux, Windows et Mac
(en théorie)

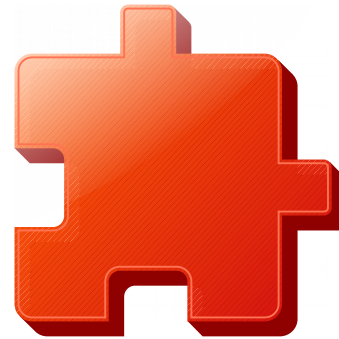
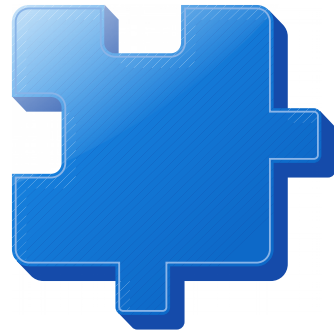


Qui a déjà servi...

Et a évité de nombreux clics douloureux !

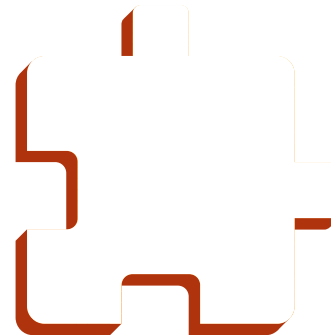
Mapping Learning : 4 briques

Lecture/Ecriture
données
(géographiques)



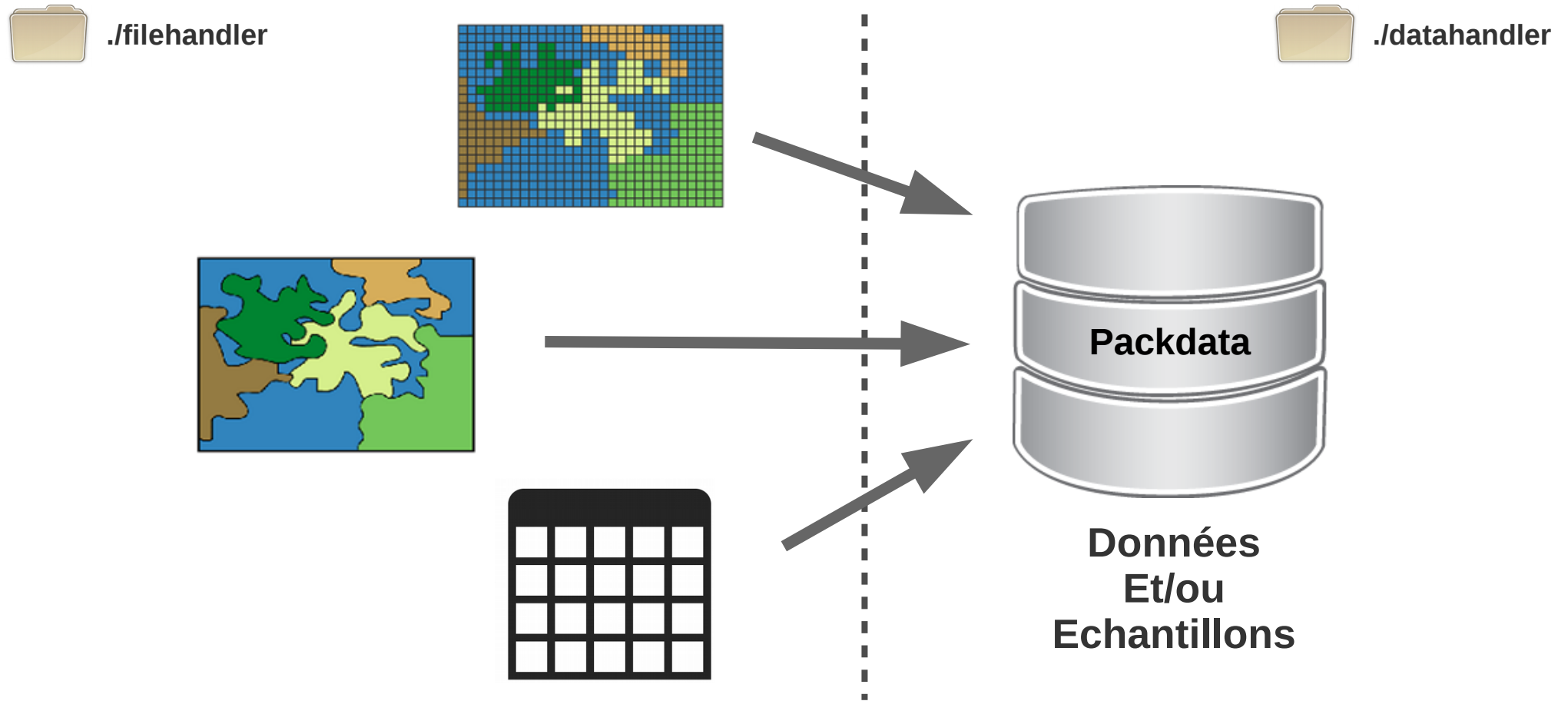
Algorithmes de
Machine learning

Application
(chaîne de
traitements)



Base de
Connaissances

Mapping Learning : Chargement des données

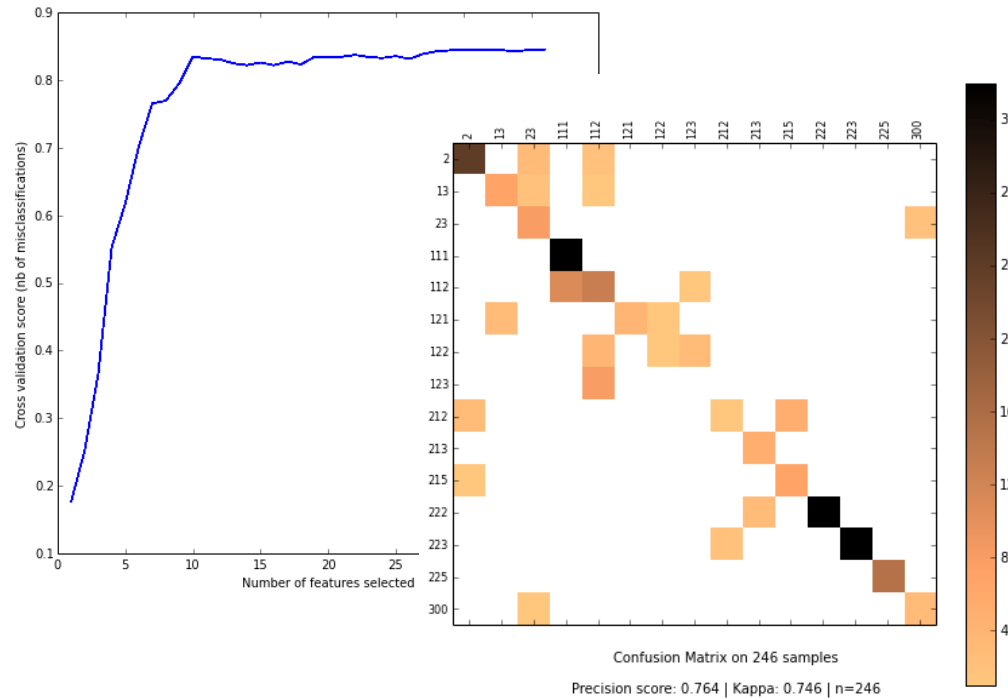


→ Utilisable en télédétection classique ou objets, ou en machine learning

Mapping Learning : (pré)traitements



Sélection de Features

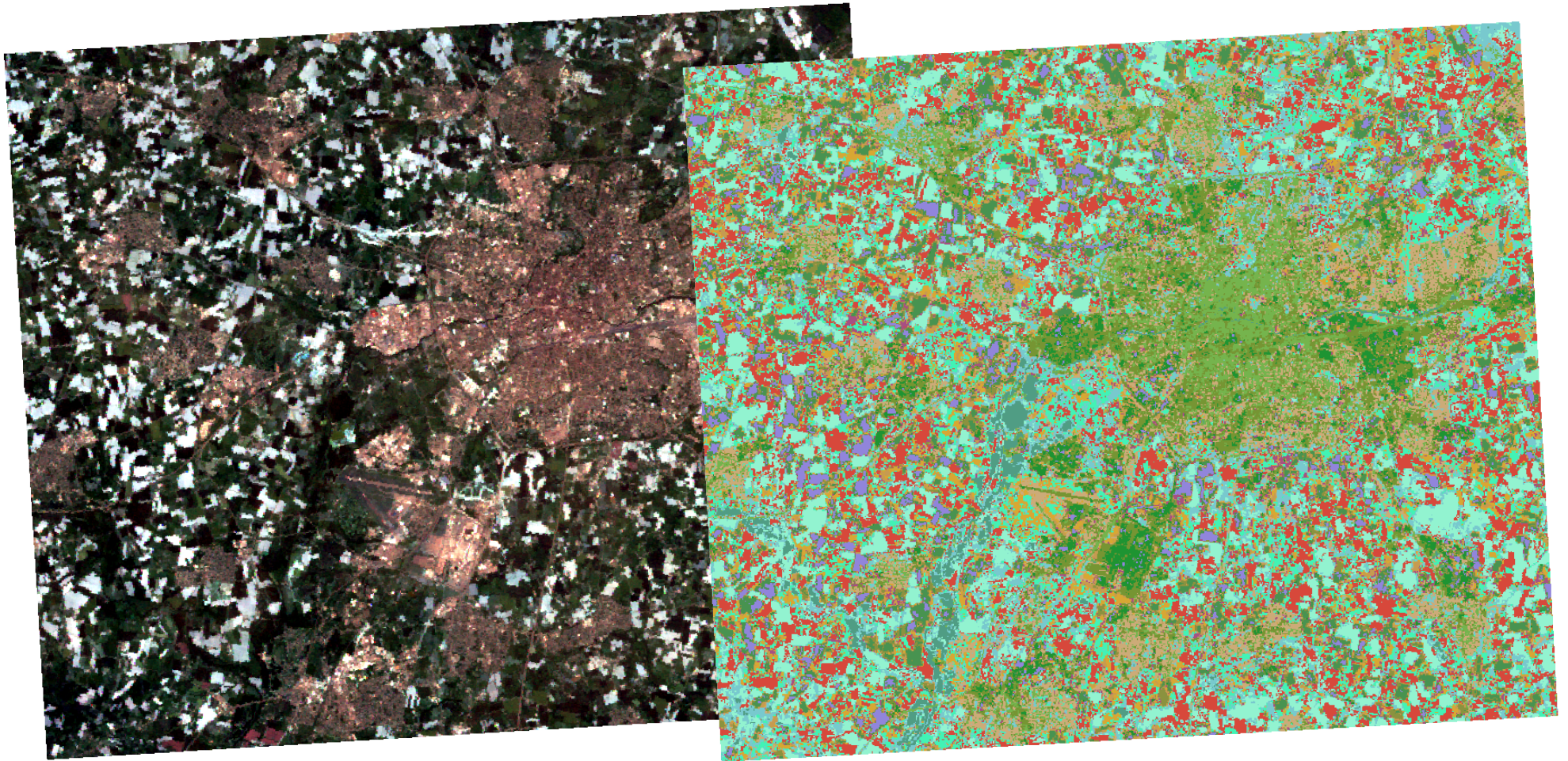


Matrice de confusion

	Précision	Recall	F1	Kappa
knn	0.99 (+/-0.01)	0.99 (+/-0.01)	0.99 (+/-0.01)	0.98 (+/-0.01)
svm	0.98 (+/-0.01)	0.98 (+/-0.02)	0.98 (+/-0.02)	0.97 (+/-0.02)
rforest	0.98 (+/-0.02)	0.97 (+/-0.02)	0.97 (+/-0.02)	0.94 (+/-0.02)

Comparaison statistique entre 3 algorithmes (3-fold)

Mapping Learning : sortie carto



Classification non supervisée d'une image satellite (Rennes)

Mapping learning n'est pas que du code

- **Aussi un projet éducatif**
- **Etudiants de**
 - Université Rennes 2
 - Université Rennes 1
 - INRIA
- **Contribueront en tant que**
 - Développeurs
 - Contributeurs (nourrir la base de connaissances)
 - Utilisateurs



Et vous êtes les bienvenu(e)s !

Mapping learning : roadmap

- Lecture/Ecriture données géographiques ✓ (90%)
- Prétraitements ✓ (?)
- Utilisation et comparaison des algorithmes de machine learning ✓
- Traitements ≈
 - Classification ✓
 - Clustering ≈ (résultats statistiques à ajouter)
 - Régression ≈ (50 %)
- Optimisation algorithmes ✓ (mais à alimenter par connaissances)
- Base de connaissances ✗
- Interfaces
 - Fichier de configuration ✓
 - CLI ✗ (à peine) commencée
 - GUI ✗

Mapping Learning : sprint !

2 jours intenses et des résultats :

- Faciliter installation (setup.py)
- Compatibilité Linux, Windows, Mac(je l'ai vu!)
- Python 2/Python 3
- CLI : parser de fichier de configuration qui facilitera l'utilisation de paramètres en CLI
- GUI : personne, mais une interface web (avec bdd !)
- Base de connaissances : une piste très prometteuse, avec un tout petit moteur de règles et des fichiers json

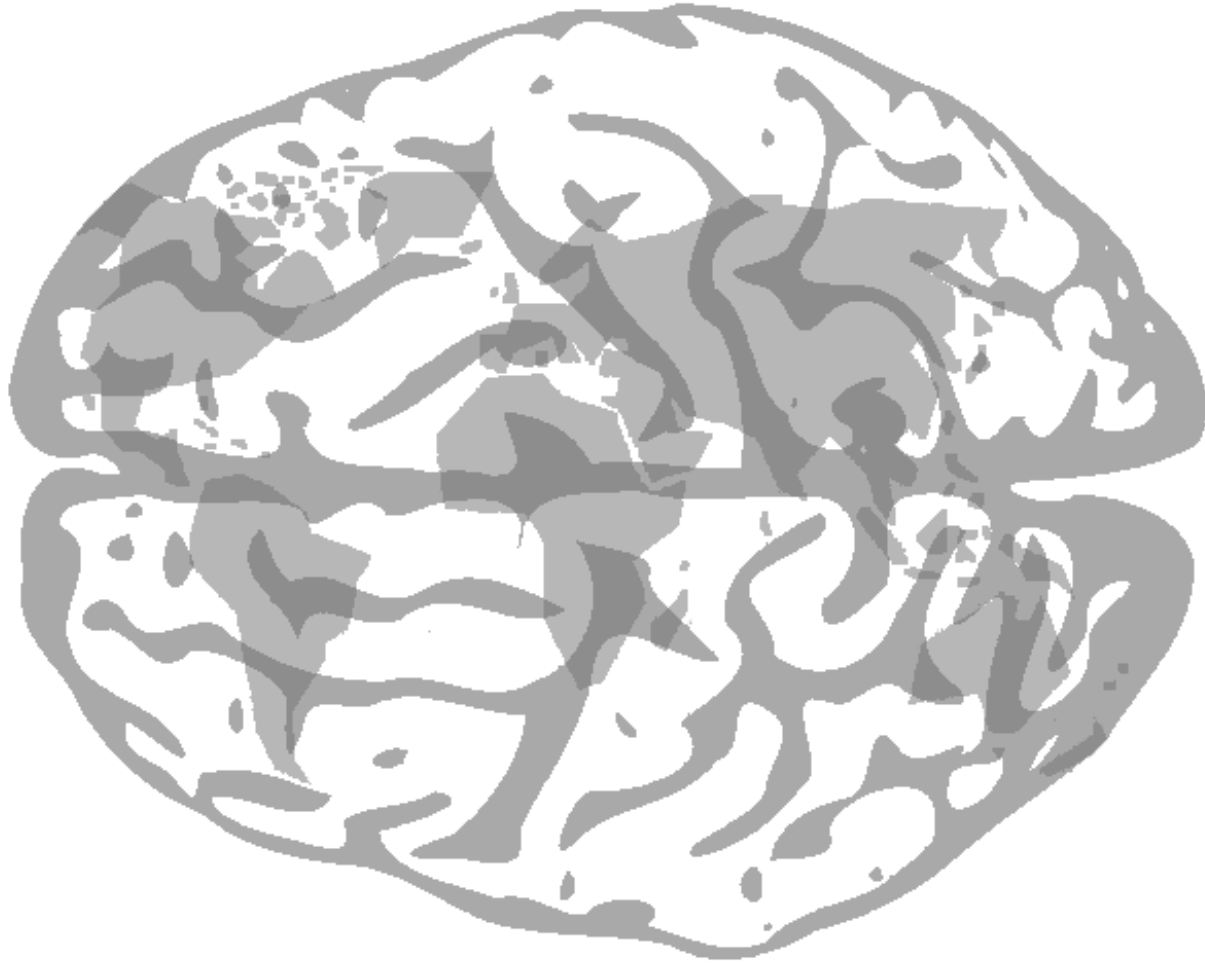
Mapping Learning : ont contribué

9 sprinters (encore merci!!!) :

- Yann Weber (programmeur au Cléo)
- Mathieu Orban (programmeur au Cléo)
- Rémi Delassus (doctorant au LaBRI / qucité)
- Côme Bertrand (programmeur chez EDF)
- Kei Delhomme (programmeur chez EDF)
- Étienne Loks (programmeur chez Iggdrasil)
- Erkang ZHANG (étudiant à CentraleSupélec)
- Hélène Artus (ingénieur télédétection)
- Nicolas Gibaud (data scientist)

**5 souhaitent continuer sur Mapping Learning !
Pourquoi pas vous ?**

Mapping learning



**Merci à l'AFPY,
aux organisateurs
de Pycon-fr, et à
Vous !**