

Programmation Fonctionnelle vs Programmation Orientée Objet : quel paradigme pour la data science ?

Sarah Diot-Girard

February 27, 2017

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

- ▶ Ingénieure-chercheuse en Machine Learning

A propos

- ▶ Ingénieure-chercheuse en Machine Learning
- ▶ Depuis juin 2016, data scientist chez WayKonet

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

**Sarah
Diot-Girard**

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

- ▶ Ingénieure-chercheuse en Machine Learning
- ▶ Depuis juin 2016, data scientist chez WayKonet
- ▶ Développement d'une infrastructure Data Science from scratch

- ▶ Ingénieure-chercheuse en Machine Learning
- ▶ Depuis juin 2016, data scientist chez WayKonet
- ▶ Développement d'une infrastructure Data Science from scratch
- ▶ On cherche un Data Engineer !

Voyez-vous la différence ?

```
import numpy as np
arr = np.arange(25, 0, -1)
arr.sort()
```

```
import numpy as np
arr = np.arange(25, 0, -1)
arr = np.sort(arr)
```

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

Programmation orientée objet

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

- ▶ Effet de bord
- ▶ Valeurs (attributs) et code au même endroit
- ▶ Architecture segmentée par composants
- ▶ Hiérarchisation via l'héritage

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

Programmation fonctionnelle pure

- ▶ Fonction au sens mathématique, entièrement définie par ses entrées/sorties
- ▶ Facilement testable
- ▶ Objets immutables

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

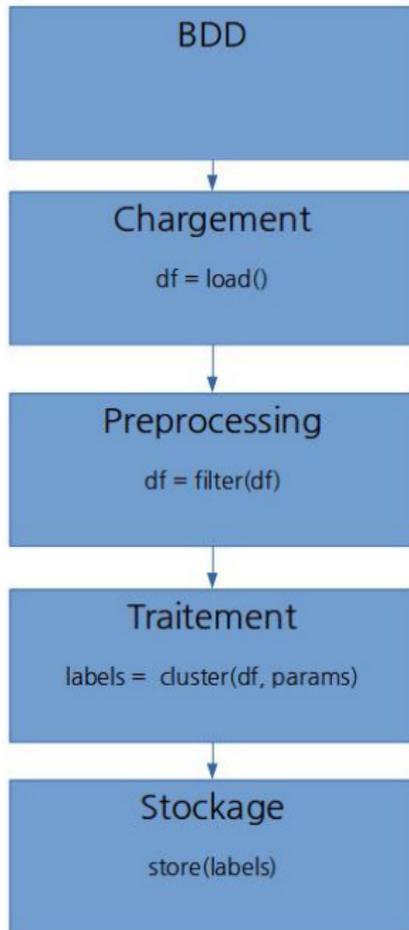
POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition



Programmation orientée objet

- + Approche scikit-learn
- + Stockage des méta-paramètres à côté de l'algorithme

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

- + Approche scikit-learn
- + Stockage des méta-paramètres à côté de l'algorithme
 - Effet de bord difficile à tester et à maîtriser
 - Parallélisation non triviale

Programmation fonctionnelle pure

- + Proche du langage mathématique
- + Correction du code

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science

**PF pour la Data
Science**

Et Python ?

Une proposition

Programmation fonctionnelle pure

- + Proche du langage mathématique
- + Correction du code
- Objets immutables -> copie

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
**PF pour la Data
Science**

Et Python ?

Une proposition

Et maintenant ?

- ▶ Python est versatile
- ▶ Numpy propose les deux versions
- ▶ Les fonctions Pandas prennent un paramètre *inplace*, mais les DataFrames sont mutables

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

Le meilleur des deux mondes ?

- ▶ Algorithmes encapsulés avec leurs méta-paramètres dans des classes
- ▶ Gestion du pipeline de données en mode fonctionnel

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

**Sarah
Diot-Girard**

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

Le meilleur des deux mondes ?

- ▶ Algorithmes encapsulés avec leurs méta-paramètres dans des classes
- ▶ Gestion du pipeline de données en mode fonctionnel
- ▶ Quid des paramètres appris ?

Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

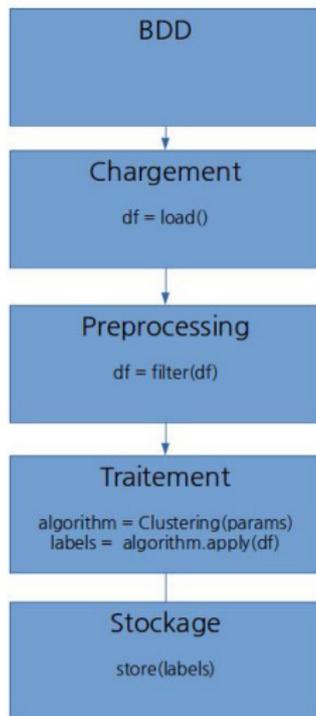
Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

Une approche mixte



Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

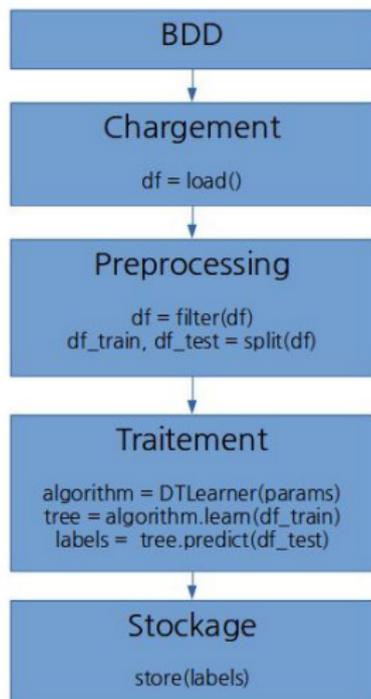
Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

Une approche mixte, avec de l'apprentissage



Programmation
Fonctionnelle vs
Programmation
Orientée Objet :
quel paradigme
pour la data
science ?

Sarah
Diot-Girard

De quoi
parle-t-on ?

POO
PF

Pipeline Data
Science

POO pour la
Data Science
PF pour la Data
Science

Et Python ?

Une proposition

Votre avis ?

De quoi
parle-t-on ?

POO

PF

Pipeline Data
Science

POO pour la
Data Science

PF pour la Data
Science

Et Python ?

Une proposition