

Majeure Machine Learning

Outils et
pipeline

Contenu



- Comparaison des langages
- Présentation des technos permettant de faire du machine learning
- Présentation de la pipeline classique d'un projet de machine learning

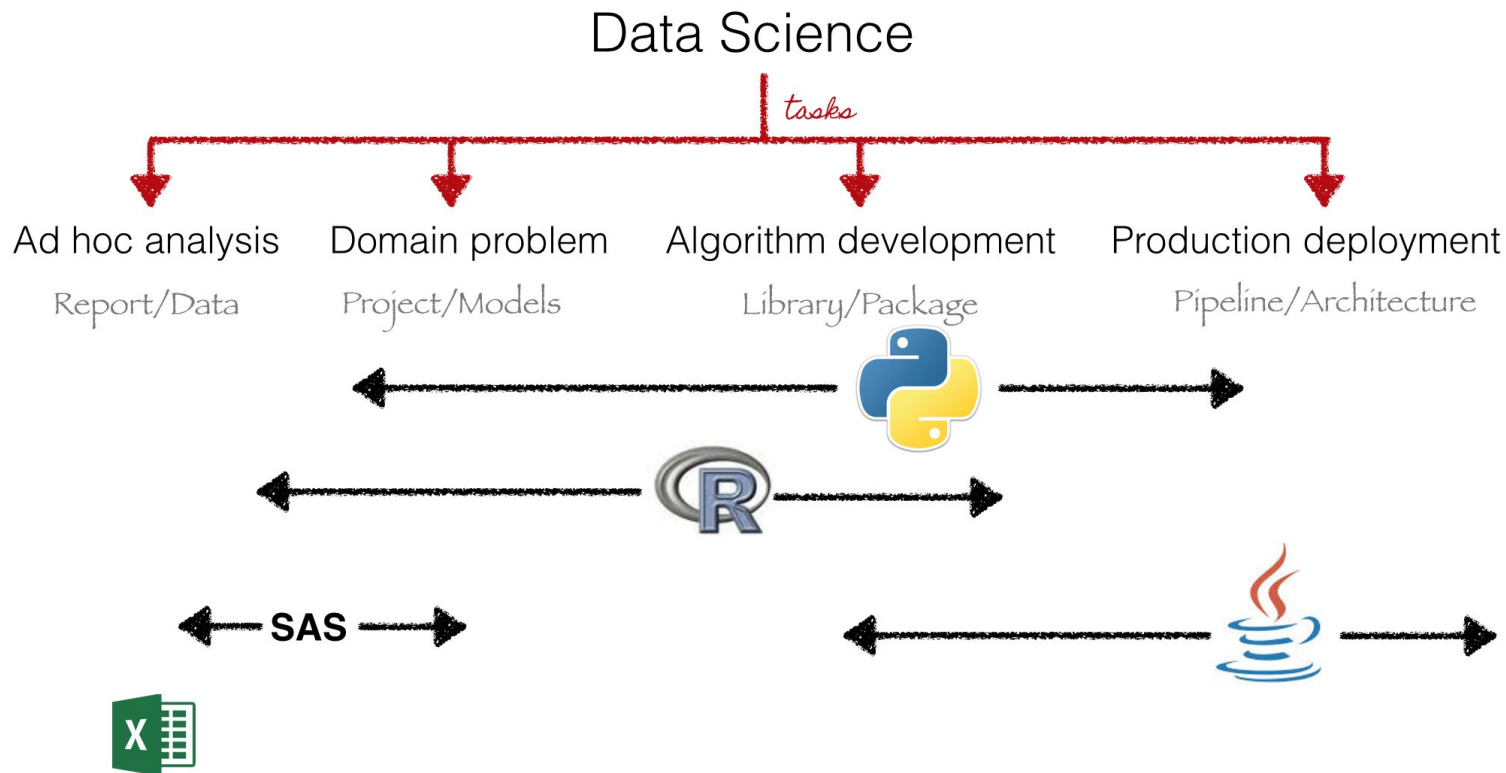
Ce que vous devrez savoir faire



- Connaître les forces et faiblesses des technos utilisées
- Etre capable de choisir les bonnes technos

Avec quels langages ?

Multi-language



R vs Python

History



Creators

Ross Ihaka and Robert Gentleman

Release Year

1995

Must Knows

1. R is an implementation of S programming language (Bell Labs).
2. R's design and evolution is handled by the R-core group and R foundation.
3. R's software environment was written primarily in C, Fortran and R.

Creator

Guido Van Rossum

Release Year

1991

Must Knows

1. Python was inspired by C, Modula-3, and particularly ABC.
2. Python gets its name from the "Monty Python's Flying Circus" comedy series.
3. Python Software Foundation (PSF) takes care of Python's advances.

Purpose

R focuses on better, user friendly data analysis, statistics and graphical models.

Python emphasizes productivity and code readability.

Used By?

R has been used primarily in academics and research. However, R is rapidly expanding into the enterprise market.

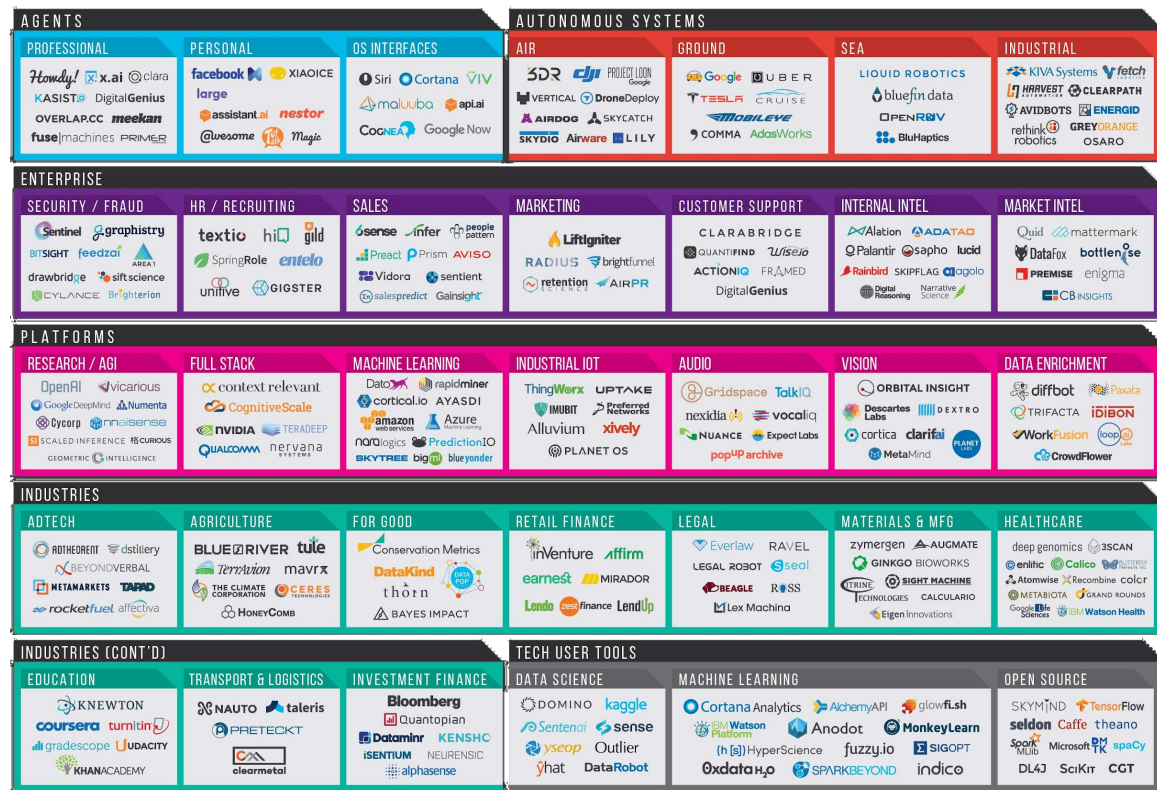
"The closer you are to statistics, research and data science, the more you might prefer R."

Python is used by programmers that want to delve into data analysis or apply statistical techniques, and by developers that turn to data science.

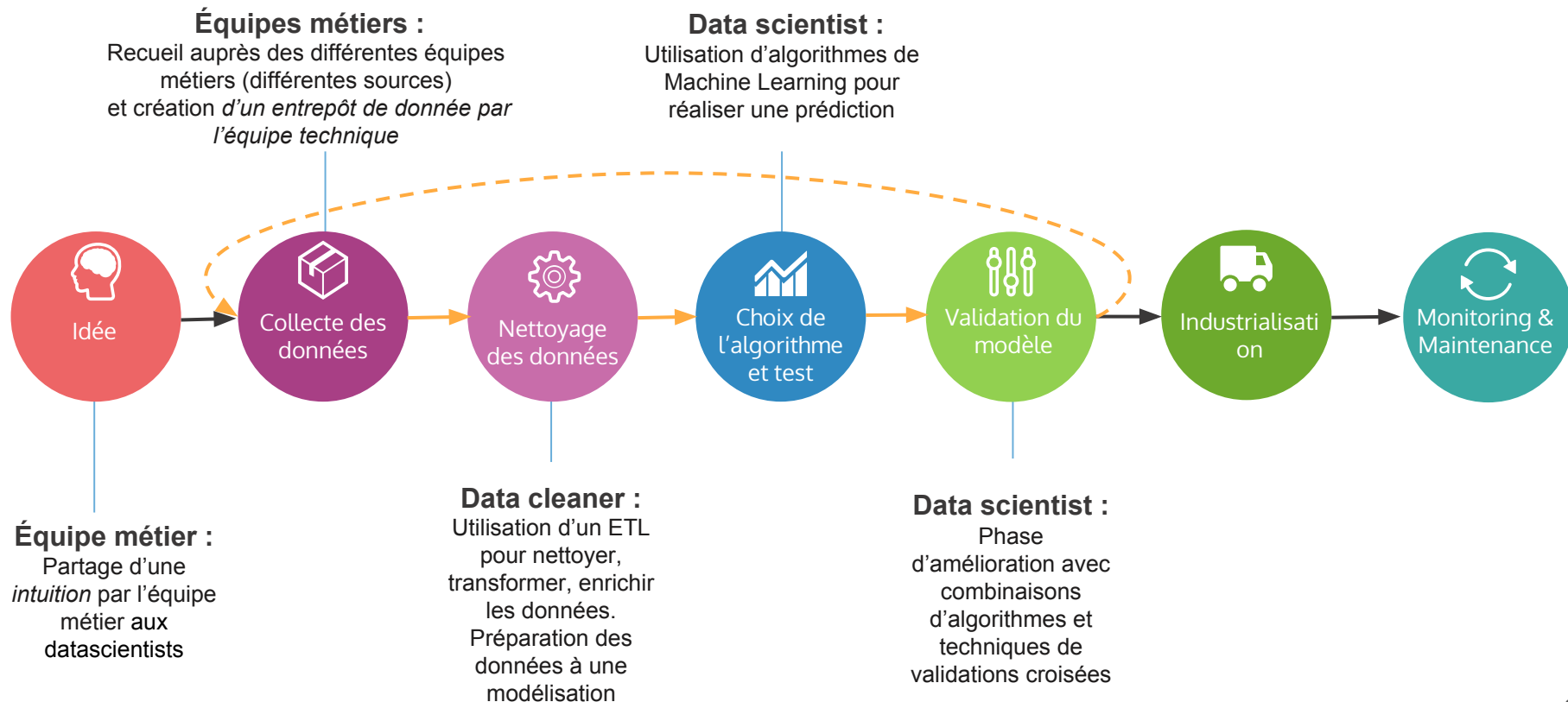
"The closer you are to working in an engineering environment, the more you might prefer Python."

La pipeline d'un projet de ML

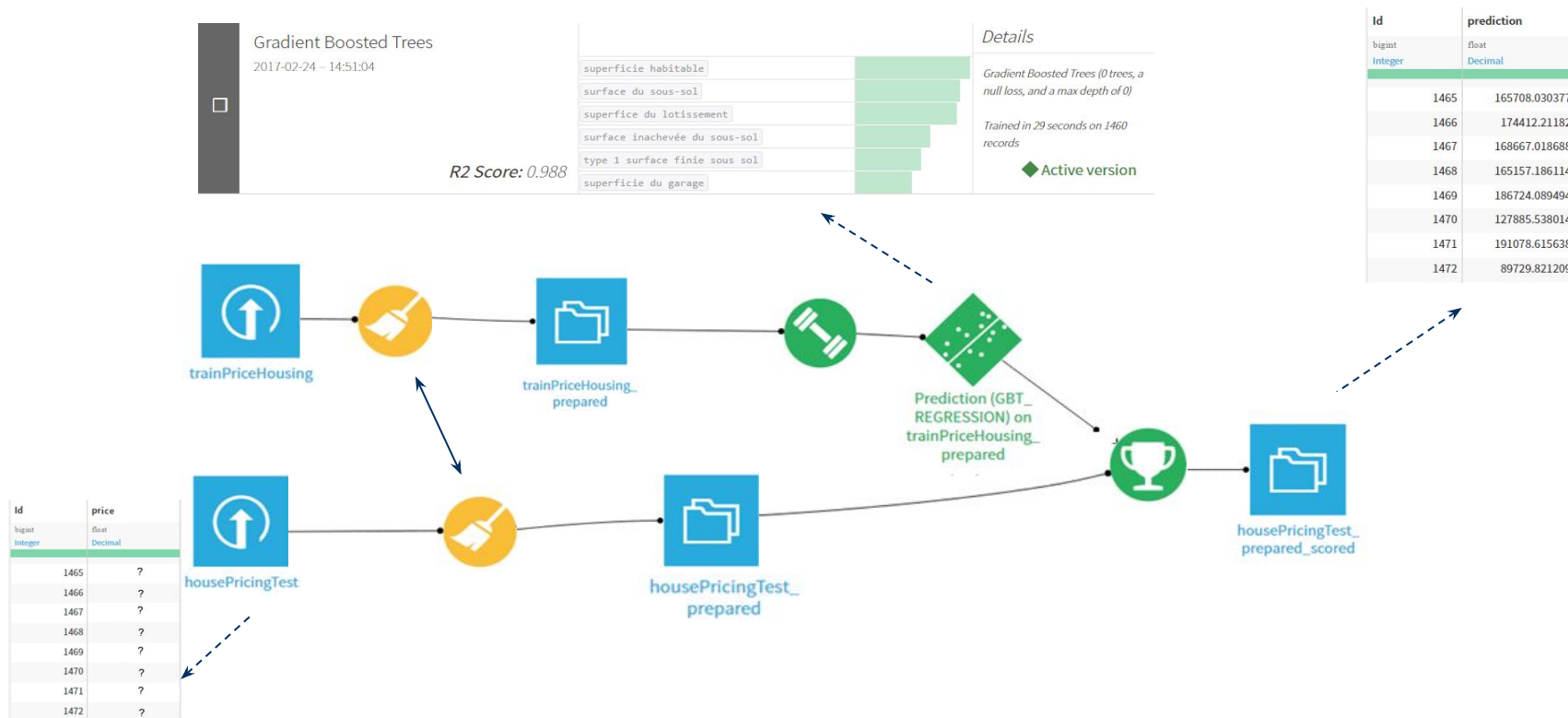
Landscape Machine Learning



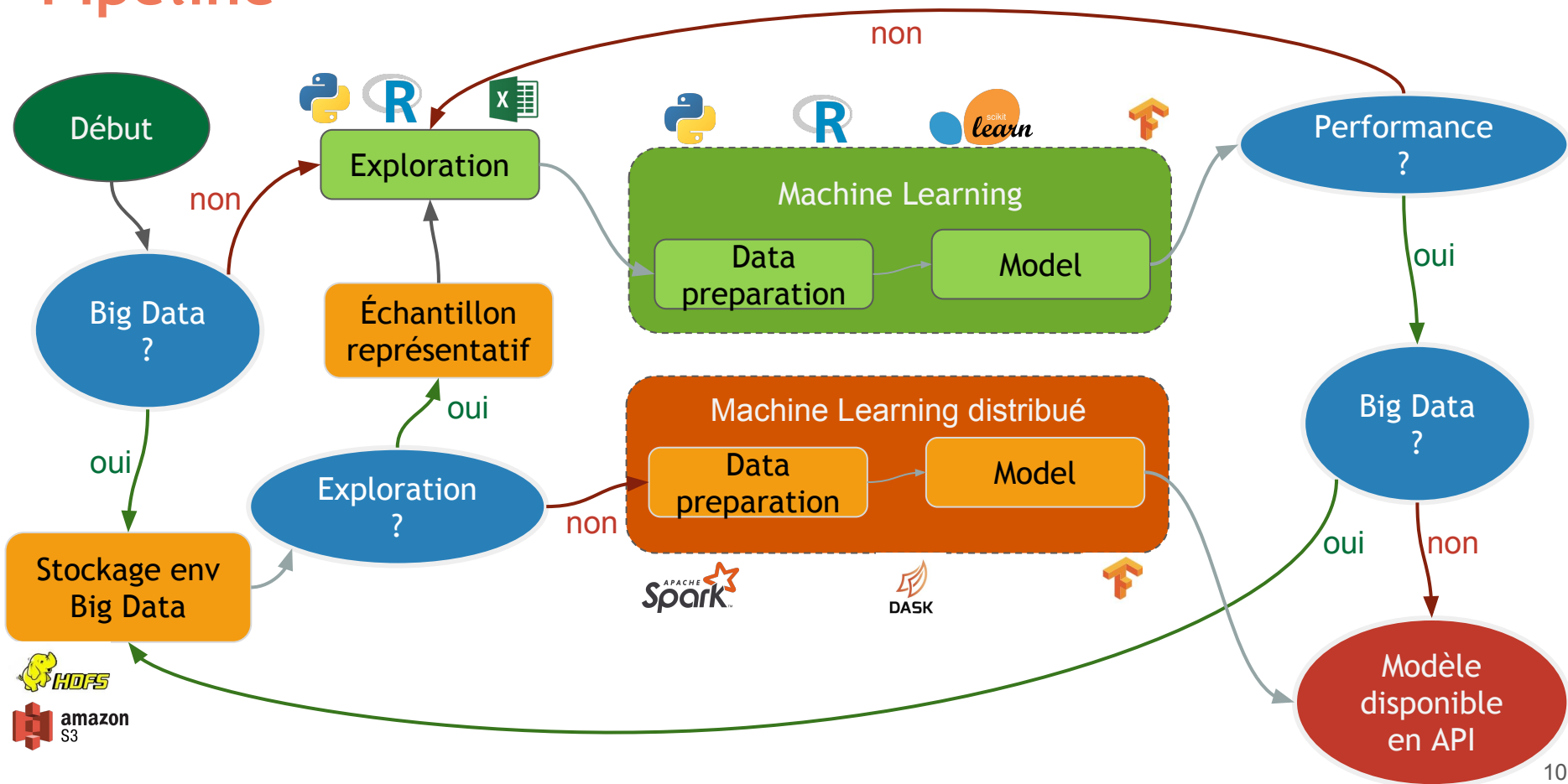
Pipeline



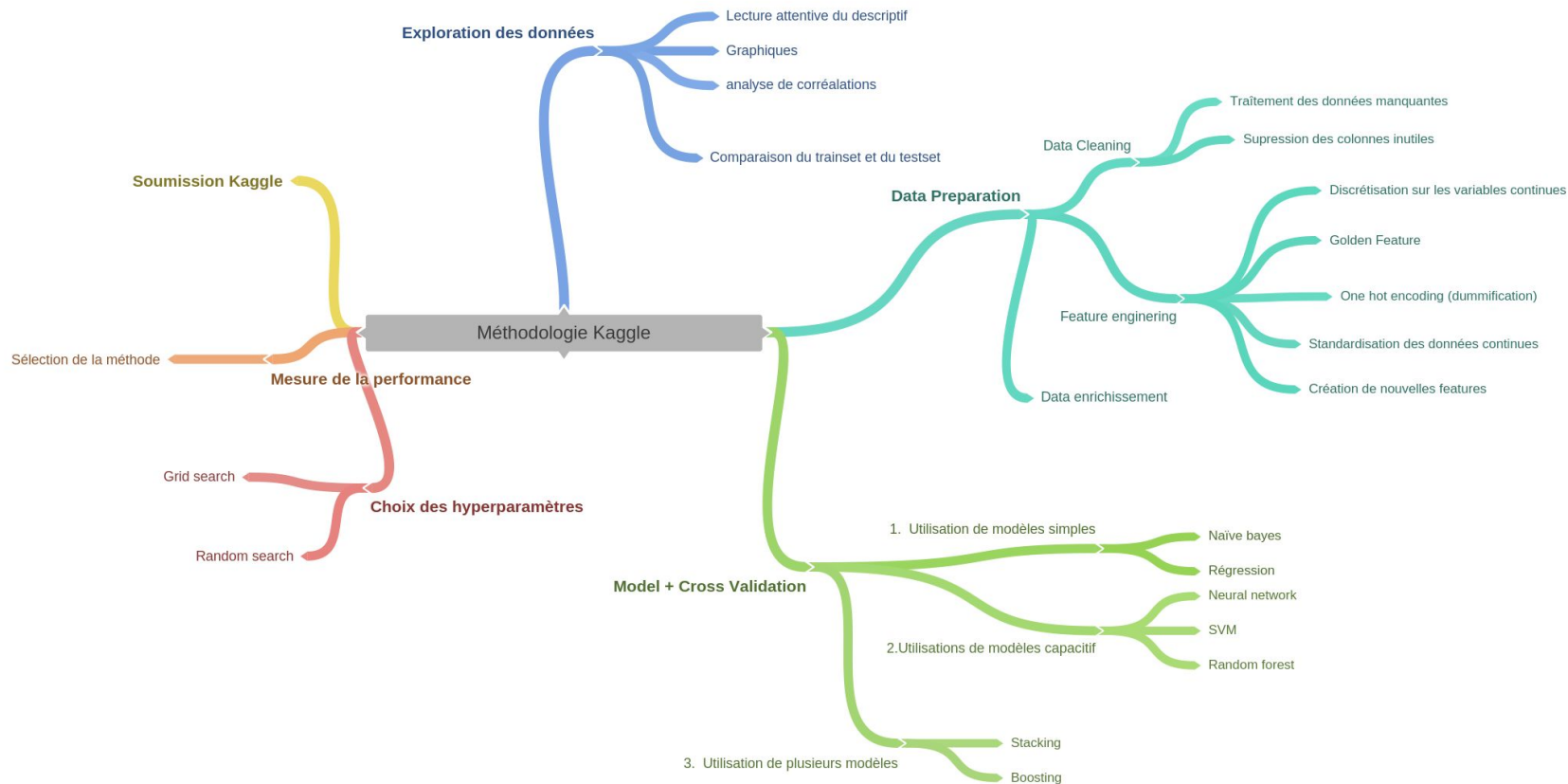
Pipeline - Dataiku



Pipeline



Pipeline kaggle





Fin du chapitre 3