

Évaluation 1

Quelques librairies et langages de programmation (hors du cours) en science de données avec descriptions des différentes caractéristiques :

1/ Langages de programmation :

Scala :

- de l'anglais Scalable language, qui signifie « langage adaptable », est un langage de programmation multi-paradigme conçu pour exprimer les modèles de programmation courants dans une forme concise et élégante
- c'est un langage modulable, c'est-à-dire qu'il permet de diviser les programmes en objets, modules et packages réutilisables
- langage fortement typé fonctionnel et orienté objet
- interopérabilité avec Java : s'intègre parfaitement avec Java et s'exécute sur la JVM (Java Virtual Machine)
- soutien au Big Data : utilisé pour le traitement de grandes quantités de données grâce à son intégration avec Apache Spark
- efficace pour manipuler les données distribuées

SQL (Structured Query Language) :

- langage déclaratif, c'est le langage standard pour interroger et manipuler des bases de données relationnelles comme MySQL, PostgreSQL et Oracle
- facilité d'apprentissage : syntaxe simple et intuitive pour les requêtes basiques, bien qu'il permette des opérations complexes comme les jointures ou les agrégations
- manipulation des données structurées : efficace pour travailler avec des données organisées en tableaux (colonnes et lignes) en permettant des opérations comme l'extraction, le filtrage, l'agrégation et la mise à jour de données
- évolutif : SQL peut s'intégrer dans des systèmes distribués pour gérer des bases de données massives.

C++ :

- langage de bas niveau performant pour des calculs intensifs et des traitements sur de grandes quantités de données.
- langage modulable, orienté objet et multi-paradigme
- utilisé pour des bibliothèques de machine learning et des systèmes critiques en temps réel
- gestion fine de la mémoire : dispose d'outils comme les pointeurs et les allocations dynamiques pour gérer manuellement les ressources.

Go (Golang)

- langage moderne conçu pour la performance, la simplicité et la scalabilité : il a en réalité une syntaxe simplifiée, minimaliste et facile à apprendre, réduisant les risques d'erreurs
- gère automatiquement la mémoire, évitant ainsi les fuites et erreurs courantes
- utilisé pour le traitement des flux de données et des pipelines d'analyse distribués
- bon choix pour des applications nécessitant une grande scalabilité, comme les plateformes de streaming de données.

Rust :

- langage moderne offrant sécurité, performance et concurrence, avec une syntaxe intuitive et puissante, combinant lisibilité et expressivité
- idéal pour la manipulation de données volumineuses et le développement d'algorithmes nécessitant une gestion fine des ressources
- sécurité mémoire : évite les erreurs courantes liées à la mémoire (comme les dépassements de tampon) grâce à son système d'emprunts et de possession.

2/ Librairies :

PyTorch	<ul style="list-style-type: none"> ✗ Bibliothèque de machine learning spécialisée dans le deep learning ✗ Prise en charge des réseaux de neurones dynamiques (dynamic computation graphs)
Plotly	<ul style="list-style-type: none"> ✗ Bibliothèque interactive pour la visualisation des données ✗ Permet de créer des graphiques dynamiques et de les intégrer dans des applications web ✗ Prend en charge Python, R, MATLAB, et Julia
ggplot2	<ul style="list-style-type: none"> ✗ Bibliothèque graphique basée sur le concept de "Grammar of Graphics" ✗ Conçue pour créer des visualisations complexes de manière simple ✗ Idéale pour explorer et communiquer les insights issus des données
caret (Classification and Regression Training)	<ul style="list-style-type: none"> ✗ Fournit des outils pour entraîner et évaluer des modèles de machine learning ✗ Simplifie la manipulation des données, la sélection de modèles et l'évaluation
D3.js (Data-Driven Documents)	<ul style="list-style-type: none"> ✗ Bibliothèque de visualisation des données pour le web ✗ Permet de créer des graphiques interactifs en manipulant directement le DOM avec des données ✗ Principalement utilisée pour des visualisations en temps réel
FluxJL	<ul style="list-style-type: none"> ✗ Bibliothèque de deep learning pour Julia ✗ Fournit une interface simple et efficace pour construire des modèles complexes ✗ Optimisée pour les calculs scientifiques
DataFramesJL	<ul style="list-style-type: none"> ✗ Équivalent à pandas en Julia ✗ Permet de manipuler des données tabulaires de manière performante et intuitive
dlib	<ul style="list-style-type: none"> ✗ Bibliothèque de machine learning performante ✗ Utilisée pour des tâches comme la reconnaissance faciale, la détection d'objets, et l'analyse d'images ✗ Conçue pour un usage en production grâce à sa vitesse
mlpack	<ul style="list-style-type: none"> ✗ Bibliothèque open-source pour le machine learning rapide ✗ Fournit des outils pour des algorithmes scalables et des modèles complexes