



Recherche documentaire

Carnet de bord

**Analyse comparative de Python et de Julia pour le
développement d'une plateforme d'apprentissage**

Nom Prénom Numéro : SERRAF Dan 3971120

M1 Informatique DAC – UE Projet Année : 2021/2021

Encadrant : MARSALA Christophe

Sommaire

- I. Introduction**
- II. Carte heuristique**
- III. Descriptif de la recherche documentaire**
- IV. Bibliographie**
- V. Evaluation des sources**

Introduction

Dans ce projet, je cherche à réaliser une analyse comparative entre 2 langages de programmation dans le domaine du Machine Learning. Je comparerai d'une part le Python qui est largement utilisé par tous depuis plus de 30 ans et permet de coder des applications Web jusqu'à des systèmes d'IA avancé. D'autre part, nous avons le Julia qui est un récent langage de haut niveau, performant et dynamique pour le calcul scientifique.

Je commencerai par réaliser une comparaison sur l'aspect bibliographique pour mettre en évidence les librairies d'apprentissage automatique existantes pour ces 2 langages.

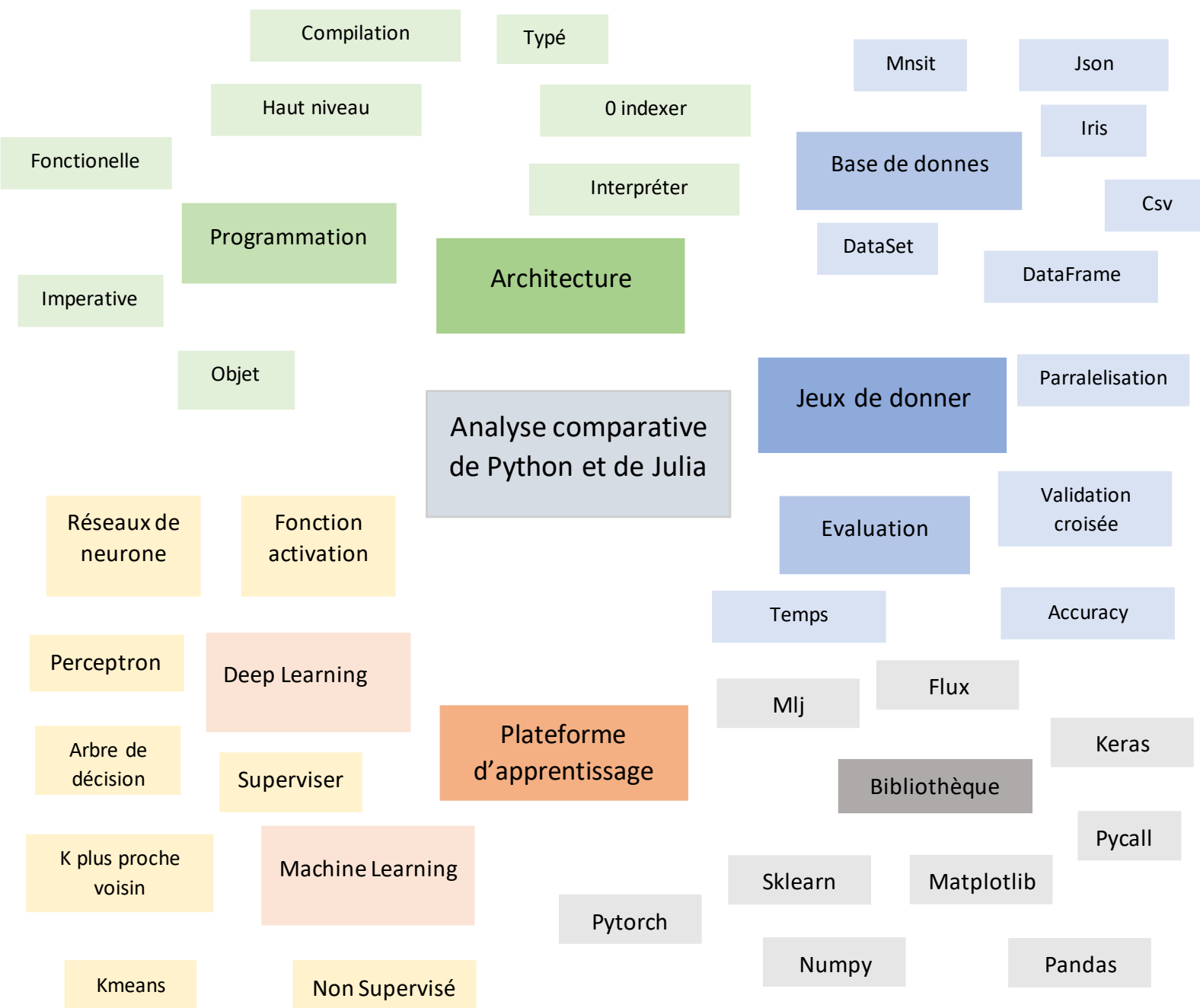
Puis je ferai une recherche sur les aspects liés à la programmation en termes de conception de solutions qui prendra en compte le style de code et l'architecture générale du programme.

Enfin, je vais créer des librairies pour l'apprentissage automatique pour chacun d'eux, je comparerai les performances en temps d'exécution de différents algorithmes de Machine Learning sur des jeux de données d'apprentissage.

Carte heuristique

Liste des mots-clés utilisés pour la recherche bibliographique :

Python, Julia, Intelligence artificielle, Apprentissage automatique, Plateforme de développement, Machine Learning, Deep Learning, réseaux de neurones , perceptron ,fonction d'activation, algorithme supervisé, algorithme non supervisé, K plus proche voisins, Kmeans, Arbre de décision,base de données, csv , mnsit, iris, json, dataframe Librairie, sklearn, matplotlib ,pycall, numpy, pandas, performance, Typage, Compilation, Structure de donnée, Programmation objet, 0-indexer,parallélisme,Jeux de donnée, Data science, benchmark, Speed, Comparative, validation croisée ,accuracy.



Descriptif de la recherche documentaire

En ce qui concerne mes recherches, j'ai d'abord utilisé le moteur de recherche Google pour trouver des sites web qui traitent des derniers sujets d'actualité en data science.

Ces derniers discutent notamment de l'aspect bibliographique des comparaisons entre Python et Julia en termes de librairie existante pour chacun d'eux pour les différents domaines du Machine Learning. Ils m'ont aussi permis de réaliser un benchmark entre eux sur les différents aspects de programmation en termes de conception de solutions et d'architecture générale tel que les structures de données, le typage ou bien encore la compilation. Ce n'est qu'après que j'ai utilisé les différents outils de recherche (tel que web of science, ACM) pour trouver des articles scientifiques qui traitent de la dernière partie de mon projet de recherche.

En effet, mon sujet étant trop récent, il n'existe pas d'article traitant de ces aspects. Pour cette dernière partie, j'ai commencé par utiliser google scholar pour avoir un premier aperçu des différents articles qui pourraient m'intéresser. En effet, google scholar a la particularité de regrouper l'ensemble des ressources en ligne, il m'a alors permis de débiter rapidement mes recherches. A l'aide de ces premiers articles trouvés, j'ai pu compléter les différents mots-clés qui m'ont permis d'accentuer mes recherches dans les catalogues de la bibliothèque. Il m'a également permis d'avoir accès à des articles qui m'ont donné une première idée sur les performances globales que je devrai retrouver lors de mes futures expérimentations.

Cette seconde recherche m'a permis de conclure que mon sujet ne possède pas encore d'article de recherche fiable dans les communautés francophones. J'ai donc dû traduire mes différents mots-clés en anglais, puis grâce à cette phase de traduction, j'ai pu compléter ma bibliographie avec des articles plus précis et plus pertinents.

Enfin, pour terminer mes recherches, j'ai décidé d'aller vérifier sur ACM (Association for Computing Machinery) qui est un outil de recherche spécifique à l'informatique et également arXiv qui a l'avantage de proposer également des articles de recherche qui sont très récents et m'éviteraient ainsi de passer à côté d'un article pertinent pour mon projet.

Bibliographie

[1]

S. Borağan Aruoba and Jesús Fernández-Villaverde. 2015. A comparison of programming languages in macroeconomics. *Journal of Economic Dynamics and Control* 58, (September 2015), 265–273. from <https://www.sciencedirect.com/science/article/abs/pii/S0165188915000883?via%3Dihub>

[2]

Viktor Axillus. 2020. Comparing Julia and Python : An investigation of the performance on image processing with deep neural networks and classification. Retrieved February 27, 2022 from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Comparing+Julia+and+Python+An+investigation+of+the+performance+on+image+processing+with+deep+neural+networks+and+classification+Viktor+Axillus&btnG=

[3]

Tyler A. Cabutto, Sean P. Heeney, Shaun V. Ault, Guifen Mao, and Jin Wang. 2018. An Overview of the Julia Programming Language. In *Proceedings of the 2018 International Conference on Computing and Big Data (ICCBD '18)*, Association for Computing Machinery, New York, NY, USA, 87–91. From <https://scholar.google.com/scholar?hl=en&q=Julia.+2018.+The+Julia+Programming+Language.+Available+at%3A+https%3A%2F%2Fjulia.org+Retrieved+August+16%2C+2018.>

[4]

Dan Segal (dan@seg.al). Julia Packages. Julia Packages. Retrieved February 27, 2022 from <https://juliapackages.com/>

[5]

White T. The Need for Speed: Julia Vs. Python. Retrieved from https://scholar.google.fr/scholar?hl=fr&as_sdt=0%2C5&q=the+need+for+speed+python+vs+julia&btnG=

[6]

2019. Top 10 Python Libraries For Data Science for 2022. *Simplilearn.com*. Retrieved February 27, 2022 from <https://www.simplilearn.com/top-python-libraries-for-data-science-article>

[7]

2020. Python vs Julia : quel est le meilleur langage pour la Data Science ? *Formation Data Science | DataScientest.com*. Retrieved February 27, 2022 from <https://datascientest.com/python-vs-julia-quel-est-le-meilleur-langage-pour-la-data-science>

[8]

2021. Python vs Julia : comparaison de ces langages pour la Data Science. Infogene. Retrieved February 27, 2022 from <https://www.infogene.fr/actualite-blog-expert/python-julia/>

[9]

2021. Julia Libraries | Top Julia Machine Learning Libraries. Analytics Vidhya. Retrieved February 27, 2022 from <https://www.analyticsvidhya.com/blog/2021/05/top-julia-machine-learning-libraries/>

[10]

2021. Les bibliothèques Python à utiliser pour le machine learning. Mobiskill. Retrieved February 27, 2022 from <https://mobiskill.fr/blog/conseils-emploi-tech/les-bibliotheque-python-a-utiliser-pour-le-machine-learning/>

* réalisé avec l'utilitaire zbib.org

Evaluation des sources

Les 4 sources que j'ai décidé de travailler sont les sources : [3],[5], [6] et [9].

Première analyse :

En ce qui concerne la ressource [3] pour la retrouver, j'ai d'abord été sur le site de primo de la Sorbonne, je me suis dirigé vers la base de données web of science où j'ai tapé la requête « Julia Programming Language » qui m'a redirigé vers le site spécialisé en informatique ACM qui contenait cette ressource [3].

Cette ressource a été publiée en 2018 lors d'une conférence sur l'informatique et le Big data qui a eu lieu à New York. Cet article a été pour moi très intéressant, car il m'a permis de mettre en avant les différences entre le Python et le Julia en termes de conception de solutions et d'architecture générale de code.

L'avantage de cet article est qu'il est accessible à tous. Il permet de comprendre de manière simple et précise les avantages de Julia par rapport à d'autres langages tels que le python. Cet article a été rédigé par différents professeurs d'informatique et de mathématiques de l'université de Valdosta au Etats-Unis ce qui donne à mon sens toute sa fiabilité. Nous pouvons noter également que ce document a été publié sur le site d'ACM ce qui lui apporte un peu plus de crédibilité.

Toutefois, bien que les données soient vérifiables à travers les références qu'ils ont publiées dans leurs articles, je trouve que leurs articles manquent d'une part d'une application propre à un jeu de données qui pourrait nous montrer à travers différents graphiques que par exemple le Julia permet d'avoir un gain de temps significatif par rapport à python.

Un second élément que je trouve manquant à travers cet article est le manque de renseignement en ce qui concerne les différentes bibliothèques propres à Julia. Malgré cela, je trouve que cet article est l'article qui m'a le plus aidé et qui se rapproche le plus de mon projet de recherche.

Deuxième analyse :

En ce qui concerne la ressource [5] pour la retrouver, j'ai été sûr Google Scholar où j'ai écrit la recherche suivante « speed Julia and Python » qui m'a alors redirigé sur ResearchGate disponible depuis le site de primo de la Sorbonne où j'ai pu télécharger la ressource.

Cette ressource a été publiée en 2022 par Mason White, qui est un professeur en informatique à l'université de Northcentral. Cet article a été publié sur ResearchGate qui est un site qui regroupe des articles de divers domaines rédigés par des scientifiques.

L'avantage de cet article est qu'il est d'une part récent et d'autre part accessible par tous et permet d'avoir un premier accès à des valeurs mettant en avant la rapidité de Julia par rapport à python. Toutefois, je suis en désaccord avec lui, il a jugé pertinent de comparer des temps d'exécution uniquement sur des opérations élémentaires, car il avait peur d'inclure un biais à travers la mesure de fonctions équivalentes. Mais il est tout à fait possible de réaliser des codes équivalents entre Python et Julia nous permettant alors d'élargir nos tests à des jeux de données d'apprentissage.

L'apport de sa bibliographie a permis de donner de la fiabilité à ces propos. Cet article m'a donc permis de compléter mon précédent article en apportant des premières mesures d'exécution à confirmer lors de mes propres expériences.

Troisième analyse :

En ce qui concerne les ressources [6] et [9], j'ai été sur le moteur de recherche Google et j'ai réalisé les requêtes suivantes « top bibliothèque Julia pour Machine Learning » et « top bibliothèque Python pour Machine Learning » qui m'a alors dirigé vers des sites d'actualités de data science. Le premier site est « simplilearn » qui a fait une liste exhaustive des 10 meilleures librairies python pour la data science. Le second site est « analyticsvidhya » qui a également réalisé une liste des bibliothèques Julia les plus pertinentes et qui couvre le plus de domaine possible du Machine Learning.

En ce qui concerne les bibliothèques python que j'ai trouvé à travers mes recherches, je connaissais déjà le nom des librairies, de ce fait vérifier la véracité des données fut simple. Par contre en ce qui concerne les bibliothèques énoncées dans l'article de Julia, les vérifier fut plus difficile.

En effet, le Julia étant récent d'une part, mais m'étant inconnu d'autre part, je n'avais aucune connaissance de ce dernier. De ce fait, j'ai dû vérifier dans la documentation officielle de Julia [4] qui contenait une page qui regroupait par catégorie les bibliothèques qui sont actuellement le plus populaire. Cette page m'a permis de confirmer mes recherches faites à travers cet article.

Toutefois ces 2 sites étant des sites connus dans l'actualité scientifique, on remarque rapidement que les éléments cités dans ces derniers sont bien pertinents. Ils m'ont d'ailleurs permis de réaliser un récapitulatif entre les différentes librairies existantes pour ces 2 langages de programmation.

Pour terminer, je trouve un peu dommage que ces articles ne citent pas un article de recherche qui traite des librairies existantes permettant ainsi d'appuyer leur propos.