

✓ **EL BRAK Imane**

I. Les technologies utilisées :

Commençant tout d'abord par la présentation des technologies utilisées dans ce projet :

➤ NumPy :



NumPy

NumPy (pour Numerical Python) est un outil parfait pour le calcul scientifique et la réalisation d'opérations de base et avancées avec des tableaux.

La bibliothèque offre de nombreuses fonctionnalités pratiques permettant d'effectuer des opérations sur des tableaux (n-arrays) et des matrices en Python. Elle permet de traiter des tableaux qui stockent des valeurs du même type de données et facilite l'exécution d'opérations mathématiques sur les tableaux (et leur vectorisation). En fait, la vectorisation des opérations mathématiques sur le type de tableau NumPy augmente les performances et accélère le temps d'exécution.

➤ Pandas :



Pandas est une bibliothèque créée pour aider les développeurs à travailler intuitivement avec des données « étiquetées » et « relationnelles ». Elle est basée sur deux structures de données principales : « Série » (unidimensionnelle, comme une liste Python) et « Dataframe » (bidimensionnelle, comme un tableau à plusieurs colonnes). Pandas permet de convertir des structures de données en objets DataFrame, de gérer les données manquantes et d'ajouter/supprimer des colonnes de DataFrame, d'imputer les fichiers manquants et de tracer les données avec un histogramme ou une boîte à moustache. C'est un outil indispensable pour la manipulation et la visualisation des données.

➤ Scikit-Learn



Scikit-Learn :Il s'agit d'une norme industrielle pour les projets de science des données basés en Python. Scikits est un groupe de paquets de SciPy qui ont été créés pour des fonctionnalités spécifiques – par exemple, le traitement d'images. Scikit-learn utilise les opérations mathématiques de SciPy pour exposer une interface concise aux algorithmes d'apprentissage machine les plus courants.

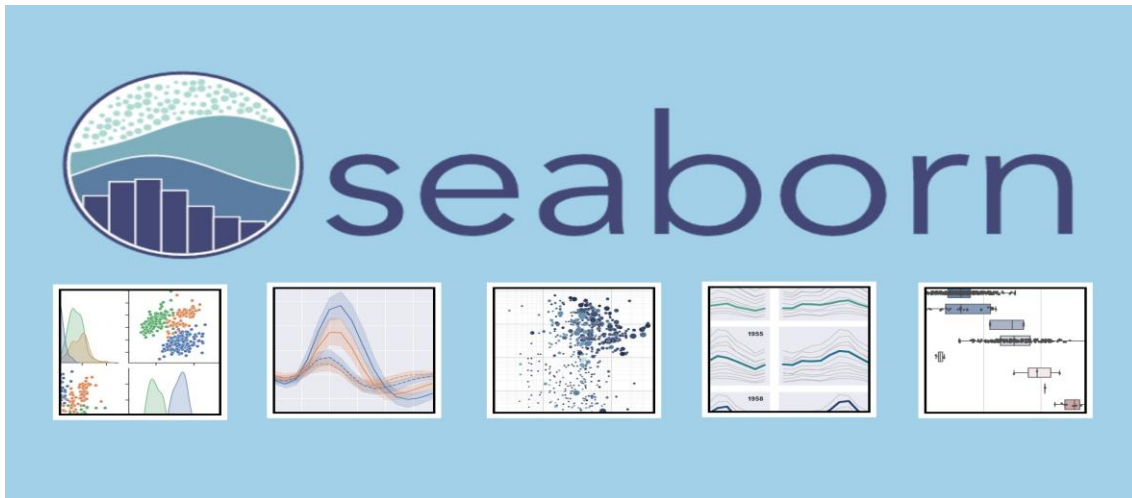
Les spécialistes des données l'utilisent pour traiter les tâches standard de Machine Learning et d'exploration de données telles que le regroupement, la régression, la sélection de modèles, la réduction de la dimensionnalité et la classification. Un autre avantage ? Il s'accompagne d'une documentation de qualité et offre des performances élevées.

➤ Matplotlib



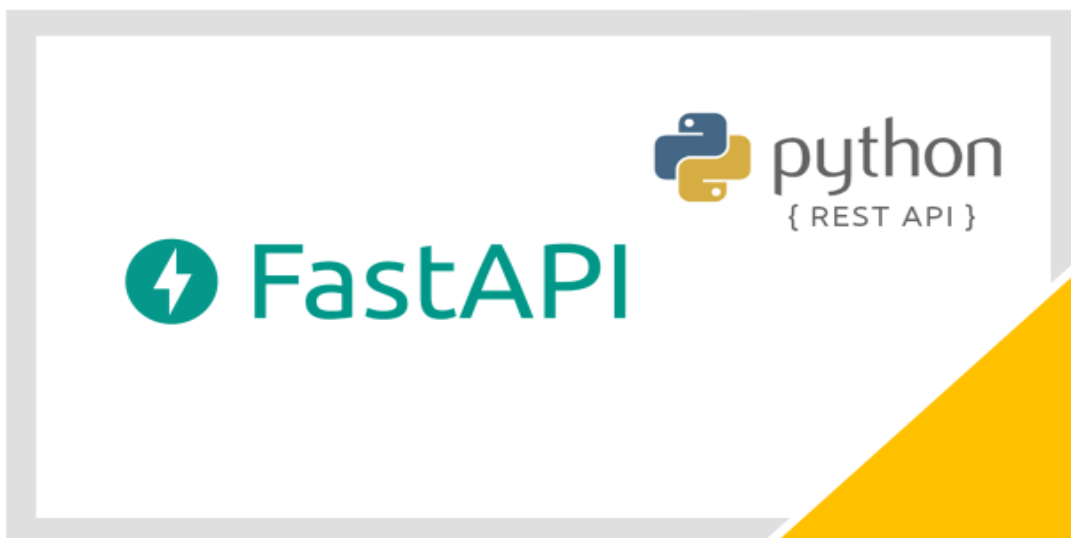
Matplotlib est une bibliothèque scientifique de données standard qui aide à générer des visualisations de données telles que des diagrammes et des graphiques bidimensionnels (histogrammes, diagrammes de dispersion, graphiques de coordonnées non cartésiennes). Matplotlib est l'une de ces bibliothèques de tracés qui sont vraiment utiles dans les projets de science des données – elle fournit une API orientée objet pour intégrer des tracés dans les applications.

➤ Seaborn :



Seaborn est basé sur Matplotlib et sert d'outil de Machine Learning Python utile pour la visualisation de modèles statistiques – cartes thermiques et autres types de visualisations qui résument les données et dépeignent les distributions globales. En utilisant cette bibliothèque, vous bénéficiez d'une vaste galerie de visualisations (y compris des visualisations complexes comme les séries temporelles, les tracés conjoints et les diagrammes de violon)

➤ FastAPI :



FastAPI est un framework web haute performance, open source, permettant de créer des APIs avec Python,
L'un des principaux cas d'usage de FastAPI est la création de endpoints API. Cette tâche peut être accomplie très facilement en utilisant le standard OpenAPI et notamment l'interface utilisateur interactive Swagger UI.

➤ Uvicorn :



Uvicorn est une implémentation de serveur Web ASGI pour Python. Jusqu'à récemment, Python manquait d'une interface serveur/application minimale de bas niveau pour les frameworks asynchrones. La spécification ASGI comble cette lacune et signifie que nous sommes désormais en mesure de commencer à créer un ensemble commun d'outils utilisables dans tous les frameworks asynchrones

Docker :



Docker est une plate-forme logicielle qui vous permet de concevoir, tester et déployer des applications rapidement. Docker intègre les logiciels dans des unités normalisées appelées conteneurs, qui rassemblent tous les éléments nécessaires à leur fonctionnement, dont les bibliothèques, les outils système, le code et l'environnement d'exécution. Avec Docker, vous pouvez facilement déployer et dimensionner des applications dans n'importe quel environnement, avec l'assurance que votre code s'exécutera correctement.

II. Kaggle :

Kaggle est une plateforme web qui accueille la plus grande communauté de Data Science au monde, avec plus de 536 000 membres actifs dans 194 pays et reçoit près de 150 000 soumissions par mois, et qui lui fournit des outils et des ressources puissants pour aider à atteindre tous les progrès de science des données.

→ Description de la compétition :



Dans cette compétition nous devons prédire les notes des joueurs AVANT qu'une partie donnée contre un bot ne soit jouée (c'est-à-dire, en fonction de leur niveau de jeu), en fonction du Game Play de Scrabble. L'algorithme d'évaluation est RMSE (Root Mean Square Error).

Le lien vers la compétition : <https://www.kaggle.com/competitions/scrabble-player-rating/overview/description>

→ Description de la dataset :

L'ensemble de données comprend des informations sur environ 73 000 jeux de Scrabble joués par trois bots sur Woogles.io : BetterBot (débutant), STEEBot (intermédiaire) et HastyBot (avancé). Les jeux sont entre les bots et leurs adversaires qui sont des utilisateurs réguliers enregistrés. Vous utilisez des métadonnées sur les jeux ainsi que les tours dans chaque jeu (c'est-à-dire, les racks des joueurs et où et ce qu'ils ont joué, AKA gameplay) pour prédire la note des adversaires humains dans l'ensemble de test (test.csv). Vous entraînerez votre modèle sur les données de jeu d'un ensemble d'adversaires humains pour faire des prédictions sur un ensemble différent d'adversaires humains dans l'ensemble de test.

Il y a des métadonnées pour chaque jeu (games.csv), des données de jeu sur les tours joués par chaque joueur dans chaque jeu (turns.csv), et les scores et notes finaux AVANT qu'un jeu donné ne soit joué pour chaque joueur dans chaque jeu (train.csv et test.csv).

Notre Score:



SOUFIANI ASSIA · 6D AGO · 7 VIEWS · PRIVATE

▲ 0

Edit



notebook69905f8abe

Python · [Scrabble Player Rating](#)

Notebook

Data

Logs

Comments (0)

Settings



Competition Notebook

[Scrabble Player Rating](#)

Run

77.4s

Private Score

206.39825

Public Score

206.61454

Best Score

206.39825 V1



Version 1 of 1

Le lien vers GitHub:

<https://github.com/Assia-Soufiani/Scrabble-player-rating>