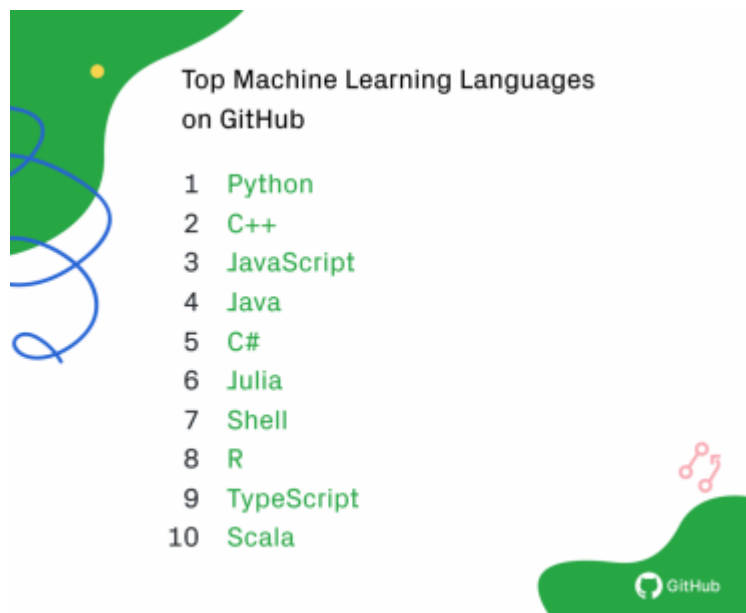


Machine Learning : qui fait quoi sur GitHub

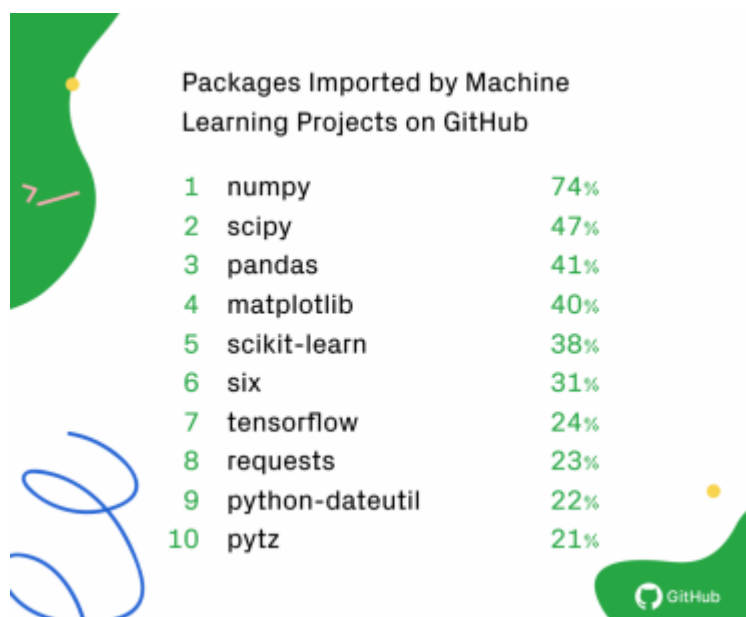
Quel est l'état du Machine Learning et la Data Science sur GitHub ? La plateforme, [propriété de Microsoft](#) depuis juin dernier, a fait un relevé des contributions sur l'ensemble de l'année 2018 en comptabilisant le code partagé, les commentaires et les [Pull Request](#)

Quel langage pour le Machine Learning ?



Python est en tête des référentiels liés au Machine Learning. A noter que les langages Julia et R sont couramment utilisés par les scientifiques des données (*data scientists*), et Scala est de plus en plus apprécié pour interagir avec des systèmes de Big Data comme Apache Spark.

Principales librairies



GitHub a calculé le pourcentage de projets ayant importé des bibliothèques Python populaires.

- [Numpy](#) prend en charge des opérations mathématiques sur des données multidimensionnelles, est la plus importée ; elle est utilisée dans près des trois quarts des projets de Machine Learning et de Data Science.

- [Scipy](#), une bibliothèque de calcul scientifique, [pandas](#), une bibliothèque de gestion de jeux de données, et [matplotlib](#), une bibliothèque de visualisations, sont utilisées dans plus de 40 % des projets Machine Learning et de Data Science.

- [Scikit-learn](#) contient des implémentations d'un grand nombre d'algorithmes de Machine Learning ; elle est utilisée dans près de 40 % des projets.

- [tensorflow](#) est un outil conçu pour travailler avec des réseaux neuronaux, est employé dans près d'un quart des bibliothèques.

Projets Machine Learning



[Tensorflow](#) est de loin le projet le plus populaire, avec plus de cinq fois plus de contributeurs que le deuxième du classement, [scikit-learn](#).

Deux projets, [explosion/spaCy](#) et [RasaHQ/rasa_nlu](#), sont consacrés aux problématiques de traitement en langage naturel (NLP), et quatre sont dédiés au traitement d'images : [CMU-Perceptual-Computing-Lab/openpose](#), [thtrieu/darkflow](#), [ageitgey/face_recognition](#) et [tesseract-ocr/tesseract](#).

Le [code source](#) du langage Julia a également enregistré l'un des plus grands nombres de contributions en 2018.