

Projet de Python

Comparaison des performances des algorithmes de tri avec
des jeux de données réels

Analyse

Dans le cadre de notre projet, nous avons privilégié trois algorithmes de tri à savoir le tri rapide, le tri fusion et le tri par tas.

- L'algorithme de tri rapide, qui choisit un élément aléatoire dans la donnée(tableau) comme pivot, le place à sa position finale, puis partitionne le tableau en deux parties : les éléments inférieurs au pivot et les éléments supérieurs au pivot.
- L'algorithme de tri fusion, qui utilise la technique de division pour régner. en triant chaque partie récursivement, puis en fusionnant les deux parties triées.
- L'algorithme de tri par tas qui effectue un tri par comparaison utilisant une structure de tas pour trier les données de manière efficace.

Sur ce, nous allons appliquer ces différents algorithmes pour trier nos différents jeux de données.

Jeux de données

Les dataset sélectionnés sont tous au format CSV.

- Premier dataset: Commercial_aviation.csv

Un dataset qui contient 899 lignes et 29 Colonnes. Après suppression radicale des données manquantes et de doublons, on se retrouve avec 602 lignes et 29 colonnes.

- Deuxième dataset: Transport.csv

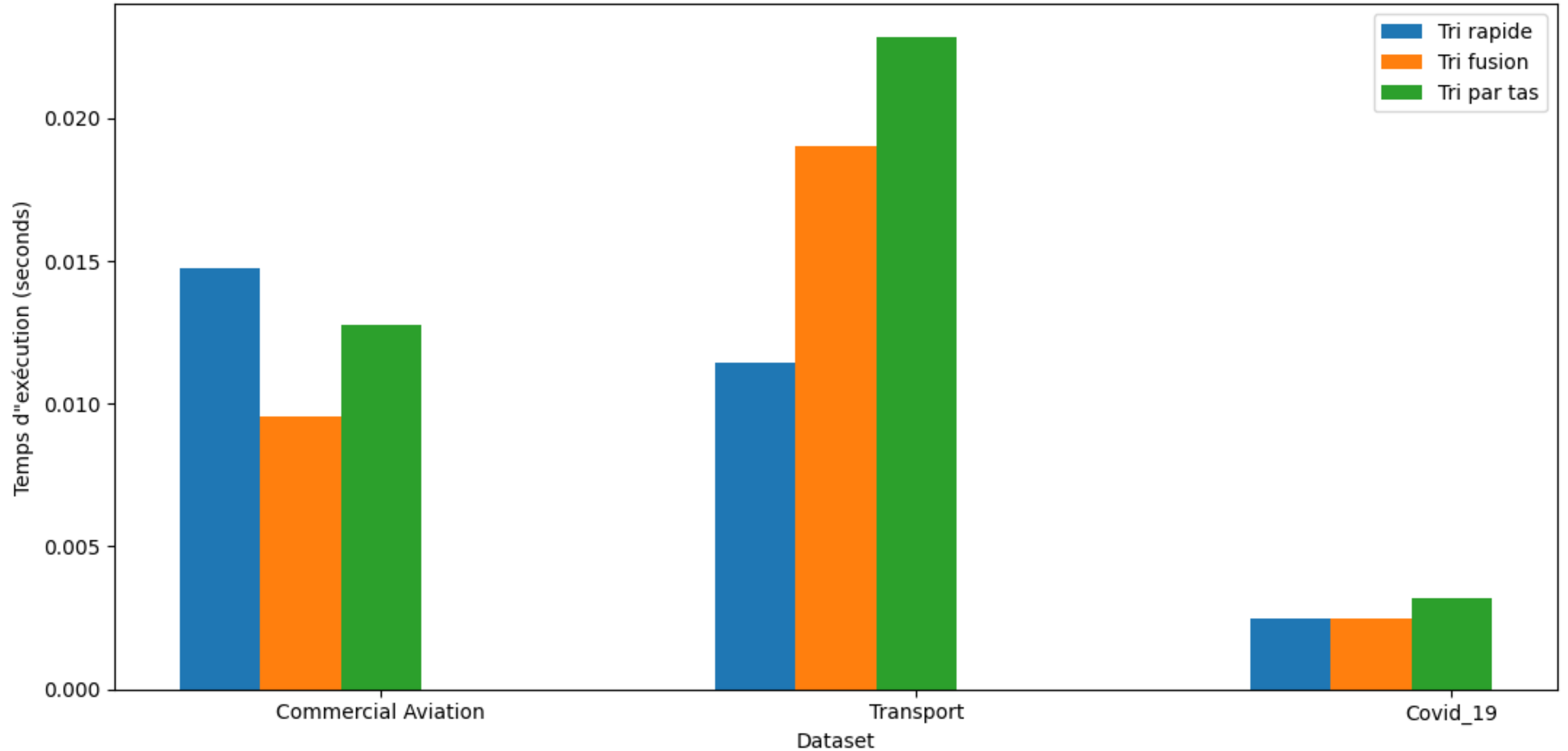
Un dataset qui contient 1884 lignes et 13 Colonnes, ne contenant pas de doublons ni de données manquantes.

- Troisième dataset: COVID-19_Outcomes_by_Vaccination.csv

Un dataset qui contient 3753 lignes et 21 Colonnes. Après suppression radicale des données manquantes et de doublons, on se retrouve avec 286 lignes et 21 colonnes.

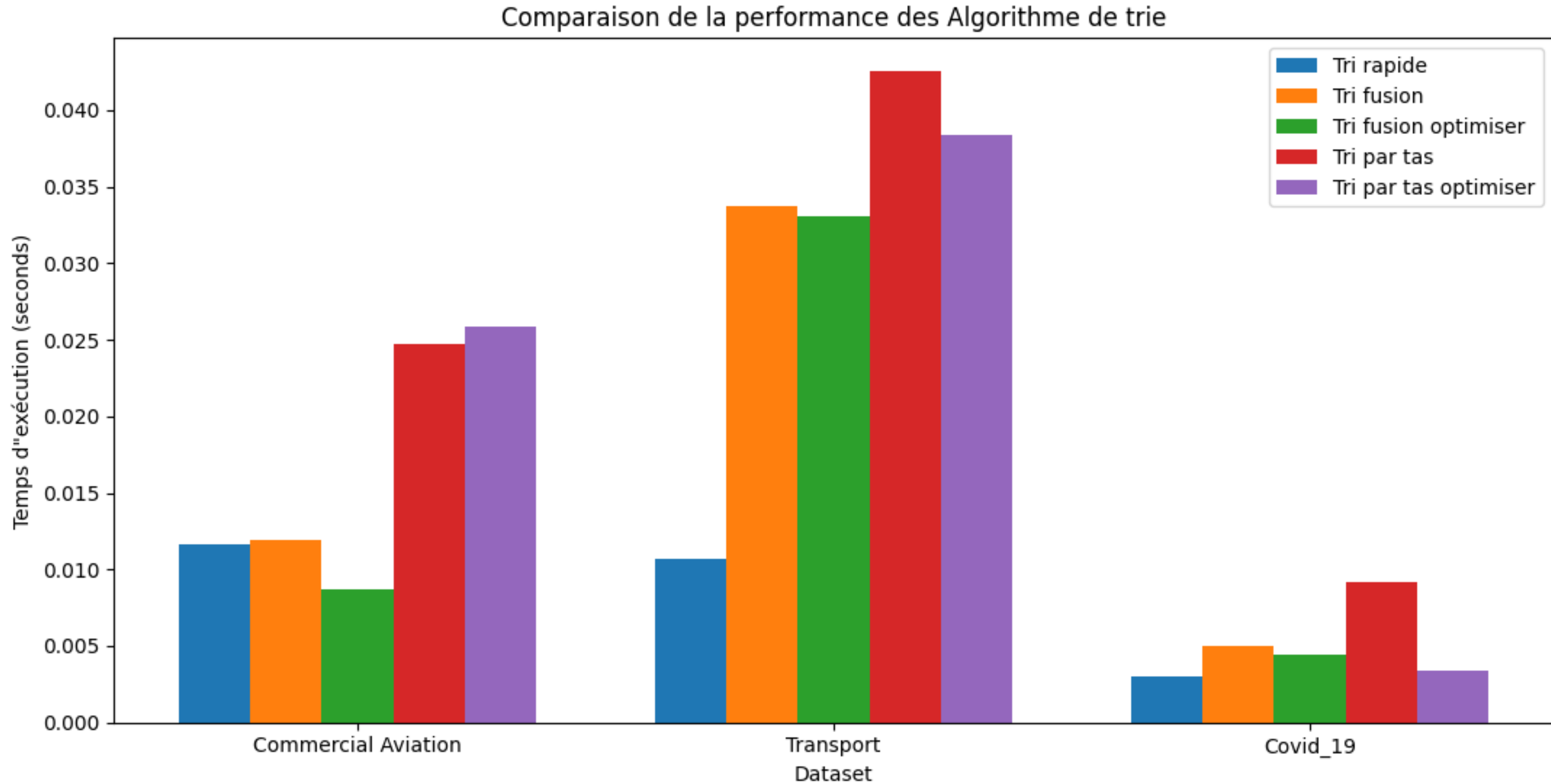
Résultat des comparaisons

Comparaison de la performance des Algorithmes de trie



A l'exécution du code et des algorithmes de tri, nous nous rendons compte que le temps d'exécution de tri diffère d'un algorithme à un autre et en fonction de la taille du dataset.

Résultat de comparaison Optimiser



Nous remarquons que les algorithmes peuvent être optimisés pour exécuter un plus rapidement les tris sur les datasets.

Nous observons un changement dans le temps d'exécution qui a considérablement diminué.

Mais pour certain dataset(Commerciale_aviation.csv), nous remarquons que le temps d'exécution du tri fusion optimiser n'a pas diminué mais à augmenter. De ce fait, nous concluons que le temps d'exécution du tri depend donc la structure de la donnée.