

Data Science 2

Blatt 11, Abgabe am 22.01.2025 um 12:00

Aufgabe 1. *Implementieren Sie parallelisiertes BFR-Clustering und betrachten Sie die gefundenen Cluster auf dem Taxi-Datensatz, den Sie zuvor exploriert haben.*

- *Schreiben Sie auf, welche Features zum Einsatz kamen und welche Datenmenge (welcher Zeitraum).*
- *Können Sie die Cluster visualisieren? Lässt sich die latente Variable 'Clusterzugehörigkeit' interpretieren?*
- *Welche unerwarteten Schwierigkeiten traten bei der Umsetzung von Parallelisierung auf (also unabhängig von den Schwierigkeiten, die eine einfache Implementierung von BFR oder der gewählte Datensatz mit sich bringen)?*