**K-anonymity**

L’algoritmo Mondrian è un’implementazione [specifica](https://www.researchgate.net/profile/Abeer-Aftab-2/publication/360426930_Demonstrating_K-Anonymity_using_Mondrian_Algorithm/links/62756863973bbb29cc67b962/Demonstrating-K-Anonymity-using-Mondrian-Algorithm.pdf?origin=publication_detail) della k-anonimizzazione che utilizza la partizione kd-tree per dividere l'intero dataset creando gruppi di dati contenenti almeno k record [sulla base della mediana di ciascun attributo e quindi esegue la generalizzazione all’interno di ciascun gruppo](https://www.researchgate.net/profile/Abeer-Aftab-2/publication/360426930_Demonstrating_K-Anonymity_using_Mondrian_Algorithm/links/62756863973bbb29cc67b962/Demonstrating-K-Anonymity-using-Mondrian-Algorithm.pdf?origin=publication_detail). [Questo algoritmo mira a evitare una generalizzazione eccessiva dei dati e a mantenere l’utilità del dataset.](https://www.researchgate.net/profile/Abeer-Aftab-2/publication/360426930_Demonstrating_K-Anonymity_using_Mondrian_Algorithm/links/62756863973bbb29cc67b962/Demonstrating-K-Anonymity-using-Mondrian-Algorithm.pdf?origin=publication_detail)  [(PDF) Demonstrating K-Anonymity using Mondrian Algorithm (researchgate.net)](https://www.researchgate.net/publication/360426930_Demonstrating_K-Anonymity_using_Mondrian_Algorithm)

Il kd-tree è una struttura gerarchica costruita partizionando i dati ricorsivamente lungo la dimensione di massima varianza. [Ad ogni iterazione viene calcolata la varianza di ogni colonna e i dati vengono suddivisi in due parti sulla colonna con massima varianza 1](https://www.vlfeat.org/overview/kdtree.html). [Questo permette di organizzare i punti in uno spazio k-dimensionale e di eseguire ricerche che coinvolgono una chiave di ricerca multidimensionale (ad esempio ricerche di intervallo e ricerche del vicino più vicino)](https://en.wikipedia.org/wiki/K-d_tree)

Nel codice, viene implementata la k-anonimizzazione utilizzando tecniche di generalizzazione e soppressione per partizionare i dati in gruppi. L’algoritmo utilizzato nel codice divide il dataframe in modo ricorsivo utilizzando la funzione split, ma non sembra rispettare esattamente la tecnica kd-tree quindi possiamo dire che il nostro codice è una variante dell'algoritmo di Mondrian.

In generale, la k-anonimizzazione è una tecnica per proteggere la privacy dei dati pubblicati rendendo difficile l’identificazione di singoli record utilizzando identificatori indiretti. La k-anonimizzazione può essere ottenuta utilizzando diverse tecniche, tra cui la generalizzazione e la soppressione dei valori degli attributi. Oltre alla k-anonimizzazione, esistono anche altre tecniche per proteggere la privacy dei dati, come l-diversity e t-closeness (utilizzate nel codice), che possono essere utilizzate insieme alla k-anonimizzazione per fornire ulteriori garanzie di privacy.