Classe Statistic.

La classe elabora statistiche su un dataset di key-value usando un NewAVLTreeMap i cui nodi contengono gli elementi key, frequency e total che rappresentano rispettivamente la chiave dell'elemento, la frequenza con la quale si presenta nel dataset e la somma dei valori per ogni elemento contenuto. Inoltre la classe presenta gli attributi occur e total che tengono conto rispettivamente di tutte le occorrenze delle chiavi e la somma dei valori per ogni chiave per favorire la complessità di occurrences() e average(). La complessità del costruttore è O(nlogk) dove n raprresenta il numero di righe nel dataset mentre k il numero di chiavi.

add(k,v)

Aggiunge la coppia k-v alla mappa. Se la chiave k è già presente all'interno della mappa si aggiornano i campi frequency e total associati a quel nodo. La complessità è O(logk) dove k è il numero di chiavi all'interno del dataset.

len()

Ritorna il numero di elementi della mappa. Complessità O(1).

occurrences()

Ritorna il numero di occorrenze degli elementi della mappa. Complessità O(1).

average()

Ritorna il valore medio degli elementi inseriti nella mappa. Complessità O(1).

percentile(j)

Calcola il j-esimo percentile, per j = 0,...,99 delle frequenze delle chiavi definito come la chiave k tale che il j% delle occorrenze nel dataset abbia chiave minore o uguale a k. Complessità O(k) dove k è il numero di chiavi all'interno del dataset

median()

Richiama il metodo percentile(j) con j = 50. Complessità O(k).

most\_frequences(j)

Ritorna una lista contenente le j chiavi più frequenti. Per l'implementazione di questo metodo è stata utilizzata una HeapPriorityQueue in quanto consente l'inserimento e la cancellazione in O(logj) con j numero di elementi nell’heap. Complessità O(klogj)