Mobile price classification

Il dataset (preso da kaggle https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification) contiene dati relativi ad alcuni cellulari. Il dataset contiene diverse feature descritte di seguito, si vuole predire il valore di range di prezzo sulla base degli attributi presenti nel dataset:

- battery_power: potenza della batteria (mAh)
- blue: ha il bluetooth oppure no (boolean)
- clock_speed: velocità del microprocessore
- dual_sim: ha il supporto dual sim oppure no (boolean)
- fc: mega pixels della telecamera frontale
- four_g: ha il 4G oppure no (boolean)
- int_memory: nemoria interna (GB)
- m_dep: Mobile Depth (cm)
- mobile_wt: peso
- n_cores: numero di core del processore
- pc: mega pixels della fotocamera esterna
- px_height: Pixel Resolution Height
- px_width: Pixel Resolution Width
- ram: RAM (MB)
- sc_h: altezza schermo (cm)
- sc_w: larghezza schermo (cm)
- talk_time: durata massima della batteria
- three_g: ha il 3G oppure no (boolean)
- touch_screen: ha il touch screen oppure no (boolean)
- wifi: ha il wifi oppure no (boolean)
- price_range: colonna target con valori 0 (costo basso), 1 (costo medio), 2 (costo alto), 3 (costo molto alto)

Dataset download:
https://bit.ly/3xoah2F

Pipeline

5. Creare una pipeline in cui gli attributi `int_memory`, `ram` e `talk_time` sono scalati in modo che abbiano media 0 e varianza 1, gli attributi `mobile_wt` e `battery_power` sono discretizzati in 5 intervalli, e **tutti gli altri attributi sono lasciati invariati**. La pipeline deve applicare il modello DecisionTree. Valutare l'accuratezza della classificazione attraverso accuracy e confusion matrix.

6. Aggiungere alla pipeline del punto 5 la funzione SelectKBest per selezionare K feature tra quelle restituite dalla pipeline del punto 5. Utilizzare la funzione di **gridSearchCV** per selezionare il valore migliore di `K`, il numero migliore di bin in cui discretizzare i valori di `mobile_wt` e `battery_power` e i valori degli iperparametri `criterion` e `min_samples_leaf` del modello DecisionTree (scegliere a piacere alcuni valori).

7. Creare una nuova pipeline che applica la decomposizione TruncatedSVD al dataset iniziale e **aggiunge le componenti** ottenute alla pipeline del punto 5. Valutare il valore migliore per il numero di componenti di TruncatedSVD tra 2, 4 e 6.