

ESERCITAZIONE 1- POLITECNICO DI TORINO

TECHNOLOGY REPLY

Dataset: <https://www.kaggle.com/datasets/iabhishekoofficial/mobile-price-classification> (modificato)

Verrà fornito agli studenti un dataset contenente informazioni riguardanti un'analisi di mercato in ambito telefonia mobile. Il dataset conterrà informazioni riguardanti caratteristiche tecniche dei dispositivi e indicazioni sulla fascia di prezzo.

Il dataset finale da trattare è contenuto nel file : "A1_dataset.csv".

La finalità dell'analisi sarà studiare le varie caratteristiche dei dispositivi, e in particolare comprendere come queste vadano ad influenzare il prezzo del dispositivo.

Inoltre, verranno implementati tre modelli di classificazione

- 'battery_power': dimensione batteria (mAh),
- 'blue': presenza bluetooth,
- 'clock_speed': microprocessore,
- 'dual_sim': support dual sim,
- 'fc': megapixel fotocamera frontale,
- 'four_g': 4G,
- 'int_memory': memoria interna (Gigabyte),
- 'm_dep': profondità (cm),
- 'mobile_wt': peso,
- 'n_cores' : numero core del processore,
- 'pc': megapixel fotocamera principale,
- 'px_height': risoluzione pixel,
- 'px_width': risoluzione pixel,
- 'ram': memoria (Megabyte),
- 'sc_h': altezza schermo (cm),
- 'sc_w': larghezza schermo (cm),
- 'talk_time': durata batteria in chiamata,
- 'three_g': 3G,
- 'touch_screen': Touch screen,
- 'wifi': wifi,
- 'price_range': range di prezzo (label)

Richieste esercitazione:

1. Caricamento dati: lettura csv (A1_dataset.csv)
2. Analisi esplorativa
 - a. Visualizza le prime 5 righe del dataset
 - b. Verifica la presenza di valori nulli
 - c. Visualizza le informazioni del dataset: lista colonne, tipologia dati, valori non nulli e utilizzo memoria
 - d. Statistiche riassuntive sui dati. Per ogni colonna numerica esamina: valore minimo, valore massimo, valore medio ...
 - e. Tramite l'utilizzo del comando *pandas.plot* visualizza le distribuzioni di tutte le features (tipologia grafico: boxplot)

- f. Controlla che il dataset sia bilanciato
3. Manipolazione dati
 - a. Sostituisci i valori nulli della colonna 'dual_sim' con il valore zero.
 - b. Trasforma le colonne 'wifi' e 'blue' con il seguente mapping:
 - Y : 1
 - N : 0
 4. Studio di correlazione
 5. Studio statistico delle features tramite l'utilizzo di librerie grafiche
 - a. Numero di dispositivi per 'price_range'
 - b. Distribuzione dei valori, di ogni colonna, in funzione del target (box plot)
 - c. Relazione tra 'px_width' e 'px_height'
 - d. Relazione tra 'fc' e 'pc'
 - e. Visualizza, tramite diagramma a barre, il numero di dispositivi per i vari valori di 'n_cores'
 - f. Numero dispositivi in base ai valori di 'four_g' e 'three_g'
 6. Creazione nuove features
 - a. Definisci la colonna 'sc_dim' definita come ('sc_w' * 'sc_h')
 - b. Definisci la colonna 'px_dim' definita come ('px_width' * 'px_height')
 - c. Definisci la colonna '3g_4g': la nuova variabile assumerà, in funzione delle colonne 'four_g' e 'three_g', i seguenti valori:
 - 0-> 3G:n 4G:n
 - 1-> 3G:s 4G:n
 - 2-> 3G:n 4G:y
 - 3-> 3G:y 4G:y
 7. Eliminazione features
 - a. Elimina le colonne: 'sc_w', 'sc_h', 'px_width', 'px_height', 'four_g' e 'three_g'
 8. Analizza la correlazione (dataset modificato)
 9. Separa la variabile di target y ('price_range') dalle altre feature (x)
 10. Normalizzazione
 - a. Normalizza il dataset (from sklearn.preprocessing import StandardScaler)
 - b. Rappresenta, tramite istogrammi, la distribuzione di tutte le feature, prima e dopo la normalizzazione
 11. Splitta X (normalizzato) e y in train e test set usando un rapporto 70:30
 12. Classificazione: tramite l'utilizzo della libreria sklearn implementa i seguenti metodi di classificazione
 - a. Decision Tree
 - b. Logistic Regression
 - c. Support Vector Machines
 13. Comparazione modelli tramite l'utilizzo della metrica di accuratezza

(opzionale)

14. LDA: applica il metodo LDA per ridurre le dimensioni del dataset
15. Classificazione con
 - a. Decision Tree
 - b. Logistic Regression
 - c. Support Vector Machines
16. Comparazione modelli tramite l'utilizzo della metrica di accuratezza