Heart Failure Prediction Dataset

Il **dataset** si trova al link https://bit.ly/3CLdwU7 (preso da kaggle - https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction) e contiene dati relativi ad alcuni pazienti:

- **Age**: età del paziente [anni]
- **Sex**: sesso del paziente [M: maschio, F: femmina]
- **ChestPainType**: tipo di dolore toracico [TA: Angina tipica, ATA: Angina atipica, NAP: Dolore non anginoso, ASY: Asintomatico]
- **RestingBP**: pressione sanguigna a riposo [mm Hg]
- **Cholesterol**: colesterolo sierico [mm/dl]
- **FastingBS**: zucchero nel sangue a digiuno [1: se FastingBS > 120 mg/dl, 0: altrimenti]
- **RestingECG**: elettrocardiogramma a riposo [Normale, ST: anomalia dell'onda T-ST, LVH: ipertrofia ventricolare sinistra]
- **MaxHR**: frequenza cardiaca massima raggiunta [valore numerico tra 60 e 202]
- **ExerciseAngina**: angina indotta da esercizio [SÌ, NO]
- **Oldpeak**: valore numerico misurato in depressione
- **ST_Slope**: pendenza del picco [Up: in salita, Flat: piatto, Down: in discesa]
- **HeartDisease**: classe da predire [1: cardiopatia, 0: normale]

Trasformazione del dataset e predizione della cardiopatia

1. A partire dal dataset originale, eliminare eventuali attributi inutili (giustificare la scelta), eliminare eventuali istanze che contengono valori nulli, trasformare opportunamente valori categorici e dividere il dataset in modo che 3/4 degli elementi siano contenuti in un nuovo dataset "**train**" e 1/4 nel dataset "**test**" preservando le proporzioni delle classi nel target.
2. Allenare il train con il modello **DecisionTree** e valutare l'accuratezza ottenuta sia sul dataset train sia sul dataset test. Confrontare i risultati ottenuti con quelli ottenuti con una predizione basata sul modello **KNeighborsClassifier** e con la predizione effettuata da un dummy classifier a scelta.
3. Confrontare l'accuratezza ottenuta nel punto precedente con l'accuratezza che si ottiene con una **10 Fold cross validation**.
4. Scalare i valori di *RestingBP*, *Cholesterol* e *MaxHR* in un intervallo tra 0 e 1 utilizzando la funzione **MinMAxScaler**.
5. Analizzare la **correlazione tra le feature** del dataset, creare un dataframe che contiene, oltre alla colonna target, le 5 feature più correlate (positivamente) al target. La predizione effettuata con DecisionTree migliora?
6. A partire dal dataset iniziale (in cui sono stati eliminati eventuali attributi inutili ed eventuali istanze con valori nulli e sono stati opportunamente trasformati i valori categorici) trovare i **valori migliori dei parametri** *criterion* e *max_depth* del classificatore DecisionTree. Come varia l'accuratezza della predizione?