

Classificazione dei Vini

progetto di Machine Learning

Dataset contenente tredici proprietà chimiche per 178 diversi vini, appartenenti a 3 classi differenti



Resoconto del progetto

EDA

- Pulizia ed analisi dei dati; fase di preprocessing

Selezione modelli e metriche

Diversi modelli di classificazione :

- SVM
- KNN
- Random Forest
- Logistic Regression

Addestramento, Validazione e Valutazione

- Report di Classificazione (Accuracy, Recall, F1)
- Grid Search e StratifiedKFold CV
- Confusion Matrix

Risultati

confronto tra i tempi di esecuzione e l'affidabilità
(misurata con F1-Score)
nei diversi modelli

- il Random Forest è significativamente più lento degli altri, sebbene si tratti di millisecondi;
- il punteggio F1 è complessivamente simile nei diversi modelli addestrati;
- il Random Forest, Log.Reg. e SVM ottengono un punteggio perfetto, etichettando correttamente tutti i vini nel test set (n=50) ma con diverse performance in validazione.
- l'Ensemble dei tre algoritmi è molto preciso, ma più lento di Log.Reg.

La scelta dovrebbe ricadere su

Logistic Regression

- ottima performance in tempi rapidi,
- elevata interpretabilità

