

Classificazione del Superamento dell'Esame

Un esempio con Albero Decisionale

Prof. Fedeli Massimo

Obiettivo del problema

L'obiettivo è costruire un modello di **Machine Learning supervisionato** capace di prevedere l'esito di uno studente:

- **Promosso**
- **Bocciato**

La previsione si basa su tre variabili osservabili:

- Ore di studio settimanali
- Percentuale di presenza alle lezioni
- Voto medio annuale

Rappresentazione dei dati

Ogni studente è rappresentato come un vettore numerico:

$$X = [\text{ore_studio}, \text{presenza}, \text{voto_medio}]$$

L'etichetta associata è binaria:

- 1 = Promosso
- 0 = Bocciato

Il dataset contiene 20 esempi, ciascuno con caratteristiche e risultato noto.

Tipo di apprendimento

Il problema rientra nell'**apprendimento supervisionato**:

- gli input sono noti
- l'output corretto è fornito durante l'addestramento

Il compito specifico è una **classificazione binaria**.

Suddivisione del dataset

I dati vengono divisi in:

- **Training set (80%)**: utilizzato per addestrare il modello
- **Test set (20%)**: utilizzato per valutare le prestazioni

Questa separazione consente di stimare la capacità di generalizzazione del modello.

Scelta del modello

Il modello utilizzato è un **Albero Decisionale**.

Motivazioni principali:

- struttura basata su regole *if–then*
- elevata interpretabilità
- adatto a dataset piccoli
- coerente con il ragionamento umano e didattico

Come funziona un albero decisionale

Un albero decisionale:

- seleziona la variabile più informativa a ogni nodo
- suddivide i dati tramite soglie
- riduce progressivamente l'impurità dei gruppi

Ogni percorso dalla radice a una foglia rappresenta una regola decisionale.

Profondità massima dell'albero

Nel modello è impostata una **profondità massima pari a 3**.

Ciò significa:

- al massimo tre decisioni consecutive
- modello semplice e leggibile
- controllo del rischio di overfitting

Perché profondità 3

Una profondità limitata:

- evita di adattarsi eccessivamente ai dati di training
- migliora la generalizzazione
- mantiene il modello coerente con il contesto educativo

Tre livelli sono sufficienti per catturare relazioni significative senza perdere interpretabilità.

Valutazione del modello

Il modello viene valutato confrontando:

- risultato reale
- risultato previsto

La metrica principale utilizzata è:

- **accuratezza** (percentuale di previsioni corrette)

Previsione su nuovi dati

Una volta addestrato, il modello può essere utilizzato per:

- stimare l'esito di nuovi studenti
- supportare decisioni didattiche
- simulare scenari ipotetici

Il processo decisionale rimane sempre spiegabile e tracciabile.

Conclusioni

L'albero decisionale rappresenta una scelta:

- tecnicamente adeguata
- didatticamente efficace
- facilmente interpretabile

La limitazione della profondità garantisce equilibrio tra semplicità, accuratezza e robustezza del modello.