

Ghost in the Cluster

Lorenzo Cappetti, Lapo Chiostrini

26/05/2025



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Obiettivi

Come e Perché?

- Analizzare e confrontare diversi algoritmi di clustering:
 - Clustering Gerarchico
 - Clustering Spettrale
- Valutare le prestazioni dei modelli su un dataset reale.
- Studiare e utilizzare **PyCaret** come strumento per:
 - Semplificare il processo di clustering
 - Visualizzare i risultati in modo immediato

Concetti principali

Tipi di Cluster e strumenti

- **PyCaret** semplifica il clustering non supervisionato, automatizza il preprocessing, la creazione dei cluster e la valutazione dei risultati.
- Il **Clustering Gerarchico** è un metodo che crea una struttura ad albero (dendrogramma) di cluster, senza dover specificare a priori il numero di gruppi.
- Il **Clustering Spettrale** è una tecnica di clustering avanzata basata sull'analisi spettrale di un grafo costruito dai dati. È particolarmente utile quando i dati non sono separabili in modo lineare (cioè, quando K-means fallisce).

Conclusioni

Punti di forza e criticità

- **Clustering Gerarchico** si è dimostrato poco adatto al nostro caso, non cogliendo bene la struttura dei dati.
- **Clustering Spettrale**, nonostante iniziali difficoltà, ha fornito buoni risultati dopo opportune modifiche: i dati si adattavano bene alla rappresentazione in forma di grafo.
- Il notebook su Kaggle includeva un'ottima spiegazione della costruzione del dataset, facilitando l'analisi.
- Tuttavia, l'assenza di un file `requirements.txt` ha causato conflitti tra librerie e rallentato l'avvio del progetto.

Ghost in the Cluster

Andiamo al Notebook!