

# Traccia tesi machine learning supervisionato e confronto tra linguaggi di programmazione

Davide Chicco

3 ottobre 2025

[davide.chicco@unimib.it](mailto:davide.chicco@unimib.it)

Obiettivo: fare un confronto tra linguaggi di programmazione open source nell'applicazione d'un algoritmo di machine learning supervisionato a dati medici.

Linguaggi di programmazione: Java, Python, C++, R, Julia

Passaggi:

1. studiare e capire cosa sia la classificazione binaria tramite machine learning supervisionato, e capire l'algoritmo regressione lineare
2. scaricare [questi cinque datasets](#) di dati di cartelle cliniche elettroniche
3. implementare in Java uno script che legga il dataset del neuroblastoma tra quelli presenti, applichi una classificazione binaria tramite regressione lineare con leave-one-out-cross-validation, e misuri i risultati tramite il coefficiente di correlazione di Matthews (MCC)
4. eseguire lo script su Microsoft Windows e Linux, e salvare il tempo d'esecuzione e i consumi energetici
5. ripetere i passaggi precedenti (3 e 4) per gli altri linguaggi di programmazione (Python, C++, R, e Julia)

Extra:

Gli scripts vanno salvati e mantenuti tramite Git e GitHub

Blischak JD, Davenport ER, Wilson G (2016) "A quick introduction to version control with Git and GitHub". PLOS Computational Biology 12(1): e1004668.

<https://doi.org/10.1371/journal.pcbi.1004668>

Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost FdV, et al. (2016) "Ten simple rules for taking advantage of Git and GitHub". PLOS

Computational Biology 12(7): e1004947. <https://doi.org/10.1371/journal.pcbi.1004947>

Come linee guida generale per l'organizzazione del progetto computazionale e per la scrittura del diario scientifico, vanno seguite le linee guida di questi due articoli:

William Stafford Noble (2009) "A quick guide to organizing computational biology projects". PLOS Computational Biology 5(7): e1000424.

<https://doi.org/10.1371/journal.pcbi.1000424>

Santiago Schnell (2015) "Ten simple rules for a computational biologist's laboratory notebook". PLOS Computational Biology 11(9): e1004385.

<https://doi.org/10.1371/journal.pcbi.1004385>

