

PROGETTO MODULO MACHINE LEARNING / APPRENDIMENTO AUTOMATICO
A.A. 2022/2023

- 1) Scegliere tre datasets da elenco: uno per classificazione, uno per regression ed uno per analisi unsupervised.
- 2) Studiare metodi di “aggiustamento” e filtraggio delle features (selezione, riduzione, completamento, ecc...) adeguato al problema da risolvere
- 3) Implementare in Python la risoluzione dei 3 problemi scegliendo uno tra i seguenti algoritmi per ogni problematica:
 - a. Classificazione
 - i. Logistic regression with gradient descent
 - ii. Naïve bayes classification
 - iii. K-NN classification
 - iv. Decision Tree classification
 - v. MLP (1 hidden layer)
 - b. Regression
 - i. Linear regression with gradient descent
 - ii. K-NN regression
 - iii. Decision Tree regression
 - c. Unsupervised
 - i. Clustering
 - ii. PCA
- 4) Effettuare una validazione di performance per ogni algoritmo e confrontare tale validazione con una equivalente eseguita utilizzando il tool WEKA (<http://www.cs.waikato.ac.nz/ml/weka>) e con le Python API scikit-learn (<http://scikit-learn.org/stable/>)
- 5) Non utilizzare lo stesso metodo per il problema di classificazione e di regression (es: se scelto k-NN per classificazione, non scegliere k-NN anche per regression).
- 6) Il punto 2) non è necessario implementarlo da zero, ma è possibile generare il dataset finale di input agli algoritmi usando le tecniche di filtraggio e analisi features del tool WEKA o di scikit-learn.
- 7) L' esame finale consisterà in una presentazione seminariale degli algoritmi implementati, organizzata in modo da presentare prima la teoria sottostante ed in seguito la dimostrazione dell'implementazione degli algoritmi scelti.

Elenco sorgenti datasets:

UCI - Machine learning Repository:

<https://archive.ics.uci.edu/ml/datasets.html>

DATA.GOV.UK

<https://data.gov.uk/data/search>

US.GOV

<https://www.data.gov/>

KAGGLE

<https://www.kaggle.com/datasets>

WORLD BANK

<https://data.worldbank.org/data-catalog>