

Progetto MOBD 2019/2020

1 Dataset

Il dataset è stato generato artificialmente e presenta 21 attributi: i primi 20 (F1-F20) rappresentano le feature mentre l'ultimo (CLASS) rappresenta la classe. Il dataset completo presenta 10000 istanze.

2 Obiettivo

L'obiettivo del progetto è risolvere un problema di classificazione multi-classe con 4 classi utilizzando algoritmi di machine learning. Il classificatore deve prevedere la classe corretta sulla base delle 20 caratteristiche di ciascun istanza.

3 Valutazione

Il progetto può essere svolto individualmente o in gruppi composti da al massimo 2 componenti. Il dataset fornito (*training_set.csv*) comprende l'80% delle istanze del dataset originale; il restante 20% verrà utilizzato per la valutazione finale del progetto. La metrica di valutazione è l'**f1-macro** sul test set [1]. Deve essere prodotta una breve relazione in cui si giustificano le scelte fatte.

4 Modalità di consegna

Al fine di poter valutare nel migliore dei modi i progetti è importante che gli script siano chiari e parzialmente commentati. Oltre alla relazione, bisogna

consegnare il codice sorgente con gli script per l'addestramento del classificatore. Il codice deve inoltre contenere una routine che ha la funzione di pre-processare, in maniera coerente con le operazioni fatte sul training set, e valutare in modo automatico le prestazioni sul test set.

5 Riferimenti

[1] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html