

1. Istruzioni e Regole

Il progetto ha lo scopo di valutare le conoscenze su tutti gli argomenti svolti durante il corso. Per ogni requisito richiesto troverete anche il relativo punteggio ottenibile. La sua presenza priva di errori comporterà il punteggio massimo, la sua mancanza il punteggio minimo (0). Eventuali errori saranno valutati di conseguenza ma non potranno mai portare a valori inferiori a zero. La somma totale dei singoli punteggi è pari a 30.

Il codice prodotto dovrà essere opportunamente commentato e opportunamente suddiviso in differenti file .py (organizzati in maniera idonea) e dovrà essere presente un file *main.py*, dal quale viene avviato il programma. La consegna del codice non suddiviso in modo appropriato o inserito tutto in un singolo (ad esempio nel file *main.py*) comporta forti penalità. Per eseguire il programma si dovrà eseguire il *file* che contiene le istruzioni del programma “principale”: `python main.py`

Si possono utilizzare tutte le librerie/framework di Python utilizzate durante il corso ma siete liberi di usarne anche delle altre. In quest’ultimo caso tale scelta dovrà essere giustificata **opportunamente** e dovreste fornire un file *requirements.yml* contenente tutte le librerie utilizzate per il progetto e pronto per la creazione del nuovo environment.

Il codice **dovrà** obbligatoriamente essere accompagnato da una relazione descrittiva di tutte le operazioni svolte e, in particolare, delle scelte fatte (es. delle tecniche di classificazione) e dei confronti (riportando tabelle, plot o altro). Tutto il codice deve essere, inoltre, obbligatoriamente commentato. La mancanza di commenti, della relazione o di una parte di essa sarà penalizzata di conseguenza.

Tutto il codice Python presentato dal docente durante il corso può essere riutilizzato per il progetto finale. **NON** sono invece ammessi progetti che presentano codice scritto da altri (studenti, siti web, etc.). Inoltre, nei casi in cui verranno consegnati progetti con porzioni di codice o relazioni palesemente simili, gli studenti coinvolti verranno convocati per sostenere una prova orale confirmatoria.

Il progetto dovrà essere sviluppato da un gruppo di max 5 studenti i cui componenti saranno estratti dai partecipanti ai singoli appelli con voto > 18 e dovrà essere consegnato ENTRO 10 giorni dalla consegna delle specifiche. La consegna anticipata del progetto ENTRO 5 giorni prevede un bonus del 10%, mentre la consegna oltre i 10 giorni prevede una penalizzazione del 10%. La consegna avverrà tramite e-mail (farà fede la data di invio) nella quale tutti i componenti del gruppo dovranno essere inclusi in CC, e tutti i file dovranno essere inclusi in un file in formato .zip.

2. Requisiti

I progetti verteranno principalmente su task di classificazione e regressione. Ogni gruppo sarà libero di affrontare il proprio dominio applicativo seguendo il proprio schema organizzativo e definendo il numero di classi e funzioni ritenute più consone. Tutti i progetti però dovranno contenere:

- [2 pt] un menù per la scelta dell’operazione da svolgere all’avvio del programma; es. selezione del classificatore/regressore, analisi del dataset, etc.
- [2 pt] analisi dei dati e della loro qualità
- [16 pt] confronto tra almeno 3 tecniche di classificazione/regressione con le metriche di valutazione delle performance più idonee, incluso il tuning degli iperparametri dove necessario
 - o [10 pt] 2 tecniche a scelta tra quelle fornite da Scikit-learn
 - o [6 pt] 1 classificatore/regressore semplice custom
- [2 pt] confronto delle precedenti prestazioni con o senza la standardizzazione dei dati
- [2 pt] confronto delle precedenti prestazioni con o senza l’uso di tecniche di aggregazione di features

- [2 pt] confronto delle precedenti prestazioni con o senza l'uso di tecniche di selezione di features
- [2 pt] confronto delle precedenti prestazioni con o senza il bilanciamento dei dati, ove necessario, o comunque inserendo opportune combinazioni di undersampling e oversampling
- [2 pt] confronto e combinazione delle precedenti tecniche al fine di individuare la combinazione ottimale (es. classificatore + tecnica di pre-processing)

NB 1 - Il tuning degli iperparametri non deve essere eseguibile dal menù, ma deve essere presente la porzione di codice con il quale è stato eseguito (eseguibile indipendentemente all'occorrenza) e la relazione deve mostrare chiaramente i passi e i risultati che hanno portato alla scelta di tali iperparametri. I classificatori selezionati tramite il menù saranno quindi da considerarsi già ottimizzati.

NB 2 – I confronti e i risultati dei confronti devono essere presentati in modo chiaro nella relazione.