



## TRACCIA PROGETTO FINALE

### Note

- **Tutti i progetti dovranno essere accompagnati da una presentazione iniziale, utilizzando o google slides o power point a piacimento (max 10 slides)**
  - Le slide dovranno essere comprensibili e chiare anche per un pubblico di non specialisti, come per esempio un cliente a digiuno di data science
  - Lo studente dovrà presentare le proprie slide in non più di 15-20 minuti
- **Ogni studente deve consegnare singolarmente il proprio progetto**
  - È severamente vietato copiare progetti altrui. Nel caso in cui lo staff di Neural riscontrerà progetti troppo simili tra loro o casi evidenti di plagio, negherà allo studente coinvolto l'attestato di partecipazione.
- **Puoi scegliere una delle seguenti opzioni per ciò che riguarda i dati a disposizione e di conseguenza il dominio che più ti piace**
- **Il procedimento strutturale da seguire è descritto nella pagina successiva**
- **Il notebook deve essere strutturato seguendo il procedimento comune richiesto. Assicuratevi che sia rigorosamente ordinato e commentato in modo da essere auto-esplicativo per chi lo leggerà**
- **La consegna e il superamento del progetto finale sono necessari per il rilascio dell'attestato del master da parte dell'accademia**
- **Il notebook, la presentazione e i dati dovranno essere consegnati attraverso l'attività 'Progetto finale' che trovate su Teams entro la mezzanotte del venerdì che precede l'ultima lezione.**

## **OPZIONE 1**

**Analizzerai un dataset sul consumo degli alcolici nei giovani.**

A [questo link](#) trovi due dataset

## **OPZIONE 2**

**Analizzerai un dataset sulla soddisfazione dei clienti di una compagnia aerea.**

Puoi scaricare il dataset da [qui](#).

## **OPZIONE 3**

**Analizzerai un dataset sulla qualità dell'acqua.**

Puoi scaricare il dataset da [qui](#).

## **OPZIONE 4**

**Analizzerai un dataset sulla salute fetale.**

Puoi scaricare il dataset da [qui](#).

## **OPZIONE 5**

**Analizzerai un dataset sul rischio del credito.**

Puoi scaricare il dataset [qui](#).

### PROCEDIMENTO COMUNE:

Inizia analizzando il dataset e le relazioni tra le sue variabili, utilizzando le tecniche di manipolazione e visualizzazione che hai appreso durante master. Commenta le tue analisi in modo chiaro.

Una volta completata l'analisi esplorativa, imposta il problema di predizione. Addestra e confronta uno o più modelli, indicando quello che ti sembra migliore. Tra i modelli che utilizzerai, includi *Random Forest* e/o *XGBoost*. Inoltre:

- per ogni modello, scegli con attenzione gli iperparametri che vuoi validare, spiegandone il significato (almeno intuitivamente)
- giustifica la scelta della/delle metriche di valutazione che scegli di adottare e analizza l'errore che commetti sul test set.

Una volta addestrati e testati i tuoi modelli, cerca di capire quali sono i fattori che influiscono maggiormente sulla predizione finale. Anche in questo caso, giustifica le scelte che fai.

### LINEE GUIDA PER LA PRESENTAZIONE:

1. Presentazione del problema, dati a disposizione e obiettivi (1-2 slides)
2. EDA (2-3 slides)
3. Caratteristiche del modello o dei modelli progettati (2-3 slides)
4. Model evaluation e feature importance (2 slides)
5. Conclusioni, insights rilevanti e sviluppi futuri (2 slides)