

Studentessa: Finiguerra Alessia

Matricola: 735326

Email: [**a.finiguerra1@studenti.uniba.it**](mailto:a.finiguerra1@studenti.uniba.it)

URL GitHub:

Immagine che contiene clipart, illustrazione, grafica, cartone animato

Il contenuto generato dall'IA potrebbe non essere corretto.

***SISTEMA DI PREDIZIONE DELLA PERSONALITA’: INTROVERSO VS EXTROVERSO***

A.A. 2024/2025

# INTRODUZIONE

La personalità di una persona è l’insieme delle caratteristiche psicologiche che influenzano il modo in cui pensa, sente e si comporta nelle diverse situazioni. Essa ci rende unici e può comprendere tratti come l’emotività, la socievolezza, la stabilità e la creatività.

Tra le molte dimensioni della personalità, una classificazione degli individui può essere quella tra persone “introverse” e persone “estroverse”: le persone introverse tendono a essere più riservate, riflessive e a preferire ambienti tranquilli, mentre quelle estroverse sono generalmente più socievoli, energiche e stimolate dal contatto con gli altri.

L’obiettivo è quello di costruire un modello in grado di prevedere accuratamente la personalità di un individuo, utilizzando le principali tecniche di apprendimento automatico.

## STRUMENTI UTILIZZATI

Il progetto è stato realizzato nel linguaggio Python utilizzando Visual Studio Code come ambiente di sviluppo. Tra le varie librerie necessarie abbiamo:

* **Pandas:** per la manipolazione e l’analisi dei dati;
* **Matplotlib e Seaborn:** per creare e gestire i grafici;

# DATA EXPLORATION E DATA PREPROCESSING

## DATASET

Per valutare e fare previsioni, un modello di machine learning ha bisogno di un set di dati da cui “imparare”, che farà quindi da base da cui costruirà la sua conoscenza.

Per predire il tipo di personalità, il nostro modello ha utilizzato il dataset “personality\_dataset.csv” (disponibile su [Kaggle](https://www.kaggle.com/datasets/rakeshkapilavai/extrovert-vs-introvert-behavior-data/data)) contenente 2900 righe e 8 colonne.

Di seguito abbiamo la spiegazione delle varie features:

* **Time\_spent\_Alone:** numero di ore che l’individuo scorre da solo ogni giorno;
* **Stage\_fear:** presenza della paura da palcoscenico;
* **Social\_event\_attendance:** frequenza di partecipazione ad eventi sociali;
* **Going\_outside:** frequenza con cui l’individuo esce;
* **Drained\_after\_socializing:** presenza dellasensazione di svuotamento dopo aver socializzato;
* **Friends\_circle\_size:** numero degli amici più intimi;
* **Post\_frequency:** frequenza di pubblicazione sui social media;
* **Personality:** questa rappresenta la variabile **target** e indica appunto se l’individuo è “Introverso” o “Estroverso”.

## VALORI DUPLICATI O NULLI

Per garantire l’integrità dei dati e prevenire un addestramento distorto del modello, è stato ripulito il dataset eliminando eventuali valori duplicati o nulli presenti al suo interno.

## EDA: EXPLORATORY DATA ANALYSIS

L’EDA (Exploratory Data Analysis) è la fase iniziale dell’analisi dei dati in cui si esplorano, visualizzano e comprendono le caratteristiche principali di un dataset.

L’obiettivo è identificare pattern, anomalie, distribuzioni e relazioni tra variabili, senza ancora applicare modelli predittivi.

Per visualizzare la frequenza con cui determinati valori compaiono nel dataset, sono stati creati dei grafici per confrontare visivamente come essi sono distribuiti rispetto alle relative caratteristiche.