Esercizio D1 – Introduzione al mondo dei dati  
  
Dataset di dati strutturati:

1. Seeds DataSet: il dataset Seeds descrive le caratteristiche fisiche di semi di grano appartenenti a tre varietà distinte. In totale, il dataset include informazioni su 7 diverse proprietà per ogni seme.Ogni seme viene etichettato con una delle tre classi corrispondenti alle tre varietà di grano, ovvero Kama, Rosa e Canadian.

Link: <https://archive.ics.uci.edu/ml/datasets/seeds>  
  
Attributi:

* Superficie
* Perimetro
* Compatezza
* Lunghezza
* Larghezza
* Cariossidi
* Coefficient di asimmetria
* Lunghezza scanalatura

Identificativi:

* Kama
* Rosa
* Canadian

1. Glass Dataset: questo dataset contiene informazioni per classificare il tipo di vetro in base alle caratteristiche chimiche ed ottiche sulla base di 214 pezzi di vetro. Ciascuno di questi pezzi contiene 10 attributi.

Link: <https://archive.ics.uci.edu/ml/datasets/glass+identification>

Attributi:

* RI: Indice di rifrazione
* Na: concentrazione di sodio
* Mg: concentrazione di magnesio
* Al: concentrazione di alluminio
* Si: concentrazione di silicio
* K: concentrazione di potassio
* Ca: concentrazione di calcio
* Ba: concentrazione di bario
* Fe: concentrazione di ferro
* Tipo di vetro:
  + Vetro per finestre di edifici trattato flottante
  + Vetro per finestre di edifici non trattato flottante
  + Vetro per veicoli trattato flottante
  + Vetro per veicoli non trattato flottante
  + Contenitori
  + Stoviglie
  + Fari anteriori

Identificativi:

* + Numero ID: un identificativo univoco per ogni esempio di vetro

1. Titanic Dataset: in questo dataset troviamo un insieme di dati che contiene informazioni sui passeggeri del tragico viaggio del Titanic nel 1912.   
     
   Link: <https://www.kaggle.com/datasets/afrahhomoud/titanic-data>

Attributi:

* Sopravvissuti ( 0 = No; 1 = Si )
* Classe passeggeri ( 1 = 1st; 2 = 2nd; 3 = 3rd )
* Nome
* Sesso
* Anni
* Numero di fratelli/sorelle a bordo
* Numero di genitori/figli a bordo
* Numero biglietto
* Tariffa
* Cabina
* Imbarco

Identificativo:  
- Passenger ID

Dataset di dati non strutturati:

1. **Recensioni prodotto Amazon**: in questo caso abbiamo dei dati testuali che possono contenere delle foto dei prodotti. Ad esempio:

"Ho comprato questo nuovo telefono e sono davvero soddisfatto! Ha una batteria che dura a lungo e una fotocamera fantastica. Lo consiglio a chiunque stia cercando un telefono affidabile e conveniente."

Questo testo non ha una struttura fissa o un formato predefinito e contiene informazioni sia sul prodotto che sulle opinioni dell'utente.

Si potrebbero fare diversi tipi di analisi, come per esempio la Sentiment Analysis per capire se si tratta di una recensione positiva o negativa; oppure si potrebbero individuare le parole chiave più utilizzate per capire meglio le esigenze dei clienti; oppure ancora si potrebbero attribuire delle categorie per le varie recensioni (es. recensioni tecniche, recensioni prezzo, recensioni funzionalità ec.).   
  
In generale l’obiettivo di queste analisi potrebbe essere quello di comprendere meglio le opinioni, i desideri e le obiezioni degli utenti su un determinato prodotto.

1. **Foto pubblicate su Instagram**: queste foto possono essere state scattate in momenti diversi, con diverse persone e in diverse situazioni, e possono contenere informazioni sulla posizione, i colori, la composizione e le emozioni degli utenti. Questi dati non hanno un formato standard o una struttura predefinita ma potrebbero essere fatte delle analisi.   
     
   Per esempio potremo assegnare delle categorie in riferimento ai luoghi in cui sono state scattate le foto oppure potremo identificare le situazioni in cui sono state scattate (es. festa, viaggio, sport, animali ecc.); oppure ancora si potrebbero identificare i rapporti tra le persone presenti nella foto (es. genitori, fratello, sorella, amici ecc.)  
     
   L'obiettivo generale di queste analisi potrebbe essere quello di comprendere meglio i profili degli utenti, i loro interessi, le loro relazioni sociali e le loro emozioni.

Studente: Federico Vacca  
Data: 01/02/2023