*GUIDA OPERATIVA PROGETTO DATA MAN E VIZ*

Al fine di una corretta riproducibilità del progetto, si fornisce la seguente guida operativa, all’interno della quale saranno presentati i notebook Jupyter e i file creati ed utilizzati per rispondere alle domande di ricerca.

In sintesi, il progetto prevede la cattura, il processamento e l’immagazzinamento dei tweet relativi alla finale di Champions League, sfruttando l’architettura Apache Kafka. Una volta immagazzinati i dati in MongoDB, vengono trattati ed arricchiti con il file di scraping relativo alla cronaca della partita. In secondo luogo, viene effettuata Sentiment Analysis su alcune collezioni di dati, per poi rispondere alle domande di ricerca attraverso una serie di visualizzazioni dei dati stessi. Di seguito, verranno presentate tutte le sottocartelle contenute nella cartella principale del progetto, nominata DATA MNGVIZ PROGETTO 2019-2020.

***CARTELLA DI RIFERIMENTO:***

* ***CODICE:***
  + **CODICE RACCOLTA E MANAGEMENT:**
    - *P\_UCLFinal.ipynb:* notebook Jupyter in cui viene impostato il producer, sfruttando Tweepy e Kafka
    - *C\_UCLFinal.ipynb:* notebook Jupyter in cui viene impostato il consumer, sfruttando Kafka e MongoDB (Pymongo)
    - *UCLFinalmanagement.*ipynb: notebook Jupyter in cui vengono trattate le collezioni immagazzinate in MongoDB, arricchite con i dati di scraping e prodotti alcuni dei dataset finali utili alle visualizzazioni.
* **CODICE SCRAPING:**
  + - *Finale\_Cronaca.ipynb:* notebook Jupyter in cui è contenuto lo scraping della cronaca della partita, presa dal sito di Repubblica
    - *ScrapingStats.ipynb:* notebook Jupyter in cui è contenuto il codice di scraping dei dati statistici della partita (utilizzati in fase di visualizzazione dei dati)
* **CODICE SENTIMENT:**
  + - *Sentiment\_ComanMbappe.ipynb:* notebook Jupyter in cui si effettua sentiment analysis relativa ai giocatori Coman e Mbappè, al fine di creare una infografica
    - *Sentiment\_Neymar.ipynb:* notebook in cui si effettuano sentiment analysis e word cloud su Neymar, al fine di creare una infografica inerente
    - *Word\_cloud\_eventi.ipynb:* word cloud inerente ad alcuni eventi particolari della partita
* **CODICE TEST VISUALIZZAZIONI:**
  + - *CorrplotVIZ:* codice R per produrre le correlation matrix per i test inerenti alle infografiche
* ***DATI:***
  + **DATI DI SCRAPING:**
    - *FinaleCronaca:* file di testo ottenuto da scraping contenente la cronaca della partita completa
    - *StatsFinale:* file di testo ottenuto da scraping contenente le statistiche della partita
* **DATI TEST VISUALIZZAZIONI:**
  + - *Test-corrplot:* file di testo contenente i dati per comporre la correlation matrix
    - *Test-barchart:* file di testo contenente i risultati ottenuti nel test psicometrico
* **DATI VISUALIZZAZIONI:**
  + *Coman-Mbappe Viz:* contiene i file Coman\_def e Mbappe\_definitivo, i file utilizzati per produrre la DataViz relativa, ottenuti dal codice Sentiment\_ComanMbappe.ipynb
  + *Neymar Viz:* contiene i file neymar\_prepartita\_fin e neymar\_postpartita\_fin, prodotti dal codice Sentiment\_Neymar.ipynb, i quali sono i risultati della wordcloud relativa. Contiene inoltre i file Ney\_pre, Ney\_post e matrix, prodotti per ottenere la DataViz relativa a Neymar
  + *Linechart Viz:* contiene i file cw1, cw2, cw3, cloud word effettuati per i tre eventi principali della partita rappresentati nel Linechart (DataViz). Per realizzare il Linechart si sfrutta anche il file “match\_int”, relativo ai dati della partita comprensiva di intervallo
  + *Introductive Viz:* contiene i file relativi alla prima infografica, ossia StatsFinale (presente anche nei file di scraping) e il file “completo”, relativo a tutti i tweet catturati nell’arco della serata, arricchiti con i dati inerenti al file di cronaca