

Analisi della relazione tra popolarità musicale e sentiment sui social media

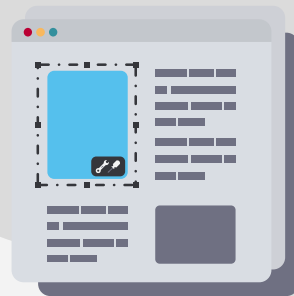
Studio sugli artisti della Billboard Hot 100 Artists
del 2023





01

Introduzione



Contesto e Obiettivi



L'avvento di piattaforme digitali e social network ha aperto nuove opportunità e sfide per gli artisti musicali:

- interazione diretta con il pubblico
- scambio di opinioni tra ascoltatori

Obiettivo dello studio: analizzare il rapporto tra popolarità musicale e percezione sociale per indagare in che misura la visibilità degli artisti si accompagni a un apprezzamento positivo da parte del pubblico.

Approccio



01

**Raccolta dati
tramite scraping**

02

Preprocessing dei dati

03

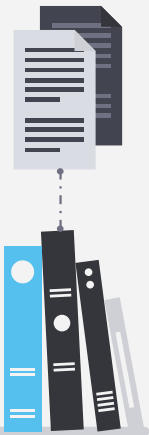
Sentiment Analysis

04

**Organizzazione dei
risultati**

05

**Rappresentazione
grafica dei risultati**



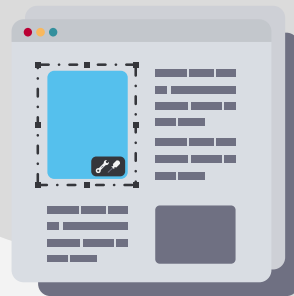


Billboard
MUSIC AWARDS

The logo features the word "Billboard" in a bold, black, sans-serif font, with the four dots of the "oo" replaced by red, yellow, blue, and green circles. Below it, "MUSIC AWARDS" is written in a similar bold, black font, underlined.

02

Tools





Strumenti Utilizzati

Selenium WebDriver: strumento open source per l'automazione dei test funzionali di applicazioni web, permette di simulare le azioni dell'utente come clic, inserimento di testo, selezione di elementi e navigazione.



VADER: strumento di sentiment analysis pensato per social media, recensioni, tweet, forum e altri testi brevi e informali.



fastText e **lid.176.bin:** utilizzati per language identification.

Matplotlib: impiegata per la creazione e la personalizzazione di grafici e visualizzazioni di dati.



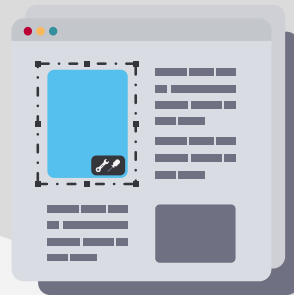


Billboard
MUSIC AWARDS

The logo features the word "Billboard" in a bold, black, sans-serif font, with the four dots of the "oo" replaced by red, yellow, blue, and green circles. Below it, "MUSIC AWARDS" is written in a similar bold, black font, underlined.

03

Raccolta dei dati





Classifica

Il primo passo è stato ottenere la lista degli artisti da considerare. È stata recuperata la classifica Billboard Hot 100 Artists del 2023 e i nomi sono stati estratti ed aggiunti a un file XML.

Le pagine Facebook e i canali Youtube sono stati aggiunti come attributi.

```
<names>
  <name youtube="morganwallen" facebook="morgancwallen" >Morgan Wallen</name>
  <name youtube="sza" facebook="sza" >SZA</name>
  <name youtube="TaylorSwift" facebook="TaylorSwift" >Taylor Swift</name>
  <name youtube="DrakeOfficial" facebook="Drake" >Drake</name>
```




Post e video

Una volta costruita la lista degli artisti con relative informazioni, il passo successivo è stato scrivere in Python due scraper, uno per Facebook e uno per Youtube.

Le pagine vengono navigate per recuperare i link ai post Facebook e ai video Youtube pubblicati.

Altri due scraper si occupano di estrarre i commenti degli utenti dalle pagine dei post e dei video.

Per quanto riguarda Youtube sono stati considerati soltanto gli ultimi 25 video che avessero almeno 20 commenti.

Per Facebook gli ultimi 15 post con almeno 15 commenti.



Commenti

I commenti estratti vengono organizzati in file XML.

Per ogni artista, in questo modo, possono esistere fino a due file, uno contenente i commenti Facebook e uno quelli Youtube.

L'elemento radice rappresenta l'artista ed ha come attributi i file considerati per l'estrazione dei commenti.

Per ogni commento è riportato il contenuto nell'attributo testo.

```
▼<Name_cardib _01="1235" _02="464" _03="795" _04="6134" _05="306" _06="125" _07="74" _08="80" _09="97" _10="91"  
_11="171" _12="573" _13="776" _14="6340" _15="2278" _16="5700" _17="1919" _18="2324" _19="3760" _20="577" _21="3732"  
_22="3080" _23="3572" _24="2540" _25="3597">  
▼<File_1.html commenti="1235">  
  <Commento_1 testo="Still the song was not a hit"/>  
  <Commento_2 testo="Raman is grazy you dont have to like you're comment only cardi b video"/>
```



Preprocessing

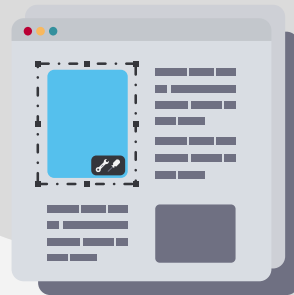
Parte del preprocessing si è svolta durante la fase di estrazione dei commenti dato che emoji, immagini e video che non fossero degli URL non sono stati inclusi nel testo.

Un'ulteriore controllo è servito soltanto per rimuovere eventuali URL presenti o eliminare commenti che non contenevano testo se non delimitatori lessicali.



04

Sentiment Analysis



Rilevazione lingua

Per ogni commento si utilizza il modello lid.176.bin con la libreria fastText per rilevare la lingua:

- si converte tutto il testo in minuscolo
- si calcolano le tre lingue più probabili considerando i primi 500 caratteri
- si salva la probabilità della lingua più probabile

Invece di procedere con l'analisi soltanto se la lingua più probabile è l'inglese, si procede anche se la probabilità che il commento sia scritto in inglese non è più piccola del 95% rispetto alla lingua più probabile.

Questo permette di analizzare anche commenti dove la probabilità che siano scritti in inglese non è la massima ma si avvicina molto.

Sentiment Analysis

Si procede con l'analisi e si salvano i risultati in due file XML, uno per i sentiment di Facebook e uno per quelli di Youtube.

I sentiment dei commenti vengono normalizzati tra -1 e 1, quindi aggregati in modo da avere un sentiment per ogni post o video, e aggregati ulteriormente per avere un sentiment per ogni artista.

```
▼<Artisti>
  ▼<Artista nome="_1997JungKook" sentiment="0.619">
    <Post nome="File_1.html" sentiment="0.798"/>
    <Post nome="File_2.html" sentiment="0.707"/>
    <Post nome="File_3.html" sentiment="0.666"/>
```

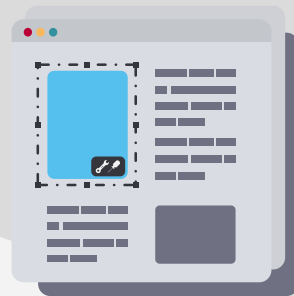


Billboard
MUSIC AWARDS

The logo features the word "Billboard" in a bold, black, sans-serif font, with the four dots of the "oo" replaced by red, yellow, blue, and green circles respectively. Below it, "MUSIC AWARDS" is written in a similar bold, black, sans-serif font, underlined.

05

Risultati



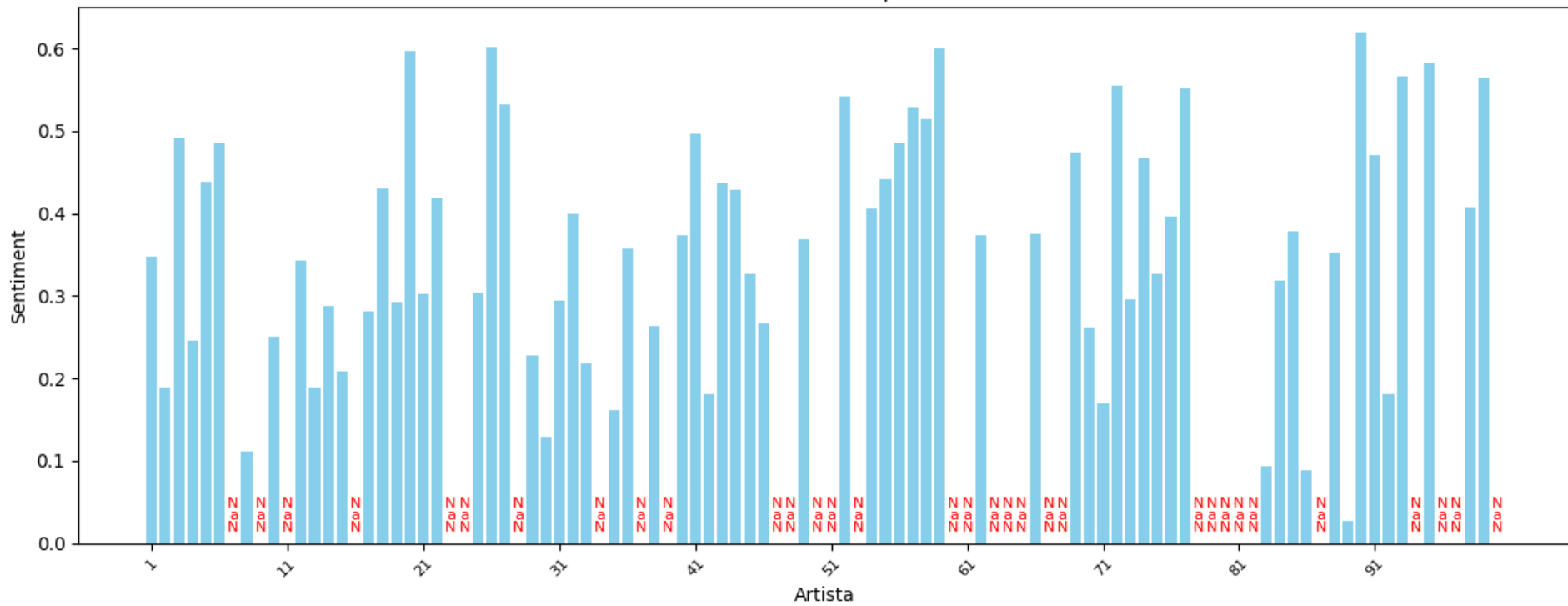
Grafici

I risultati ottenuti sono stati rappresentati sotto forma di grafico.

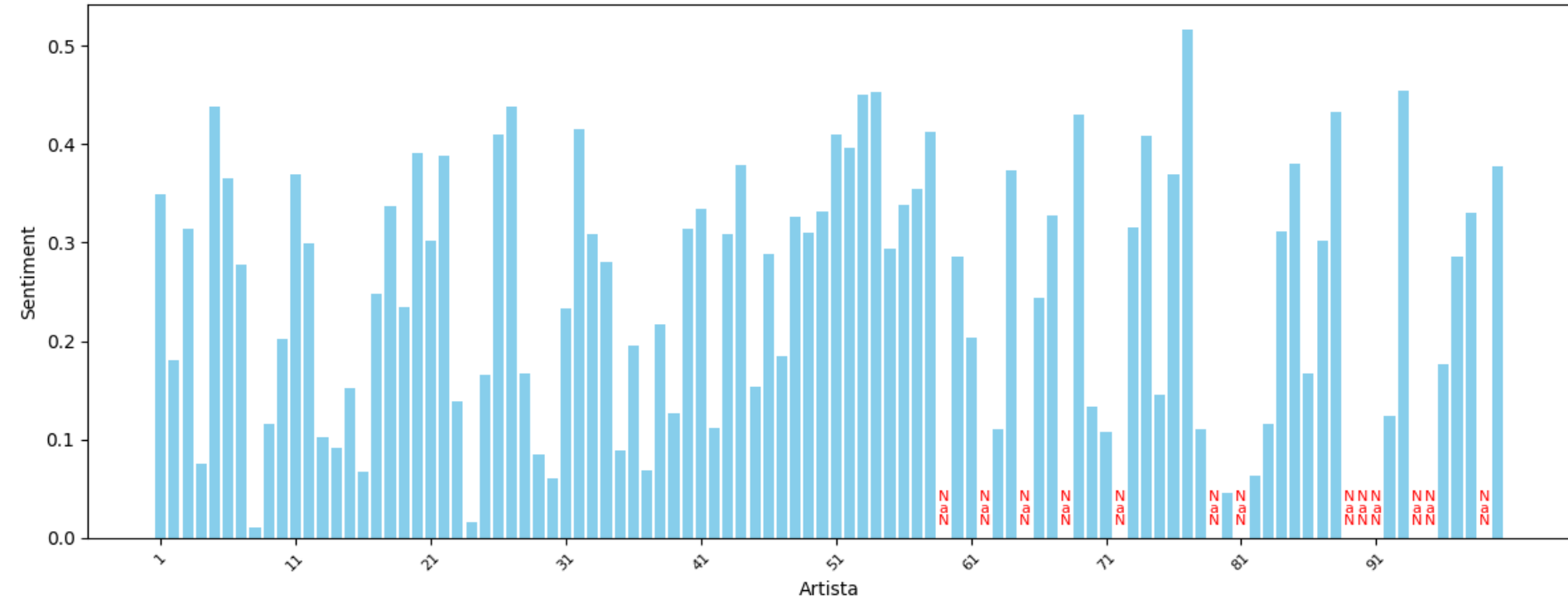
L'asse delle ascisse indica gli artisti seguendo l'ordine della classifica, quindi con una numerazione da 1 a 100, mentre quella delle ordinate indica il punteggio complessivo di sentiment, con valori tra -1 e 1.

Per gli artisti per i quali non è stato possibile recuperare la pagina Facebook o il canale Youtube ufficiali, o che non hanno superato i requisiti minimi di numero di post o video e commenti, è presente la dicitura NaN che sta ad indicare un'assenza di valore di sentiment, dato che in quei casi non è stata eseguita l'analisi.

Sentiment Facebook per artista



Sentiment Youtube per artista





Risultati

Il sentiment Facebook varia da valori compresi tra 0 e 0.05, che possono essere considerati neutri, a valori leggermente superiori lo 0.6, quindi non sono presenti sentiment negativi.

È possibile osservare che per 32 artisti il sentiment non è presente, quindi è indicato NaN.

Il sentiment Youtube in modo simile varia da valori tra 0 e 0.05, quindi neutri, a valori leggermente superiori lo 0.5.

Anche in questo caso non sono presenti sentiment negativi.

Per 13 artisti non è stato calcolato il sentiment, quindi è presente NaN.

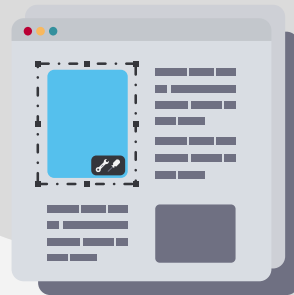


Billboard
MUSIC AWARDS

The logo features the word "Billboard" in a bold, black, sans-serif font, with the four dots of the "oo" replaced by red, yellow, blue, and green circles. Below it, "MUSIC AWARDS" is written in a similar bold, black font, underlined. The logo is set against a white rounded square background.

06

Conclusioni



Attraverso l'uso di scraper custom per Facebook e Youtube e l'applicazione della sentiment analysis con VADER, è stato possibile valutare l'apprezzamento del pubblico nei confronti degli artisti più ascoltati del 2023 secondo la classifica classifica Billboard Hot 100 Artists.

I risultati evidenziano che la maggior parte degli artisti ottiene sentiment positivo e solo una piccola parte un sentiment neutro, mentre non emergono valori negativi.

Inoltre, l'analisi mostra come non tutti gli artisti più popolari abbiano un sentiment elevato, indicando che la popolarità non coincide necessariamente con il gradimento del pubblico.

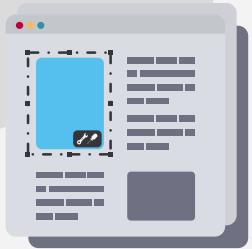
Questo suggerisce che, nel contesto musicale contemporaneo, l'ascolto e la visibilità mediatica possono essere influenzati da fattori diversi dal consenso emotivo espresso dai fan.

Sviluppi Futuri

Studi futuri potrebbero ampliare l'analisi per comprendere meglio le dinamiche di popolarità e gradimento degli artisti:

- **includere altre piattaforme social:** ogni piattaforma ha un pubblico con caratteristiche e comportamenti diversi, ampliando le piattaforme considerate si ottiene una visione più completa della popolarità e si riduce il rischio di bias.
- **analizzare il sentiment anche in lingue diverse dall'inglese:** potrebbero esserci differenze culturali e regionali che influenzano l'opinione del pubblico.
- **confrontare l'evoluzione del sentiment nel tempo:** la percezione di un artista potrebbe variare in relazione a eventi (uscita di un album, concerti, scandali, collaborazioni). Analizzare l'andamento temporale consente di capire come e perché il sentiment cambia, individuando pattern o correlazioni con fatti concreti.





**Grazie per
l'attenzione!**