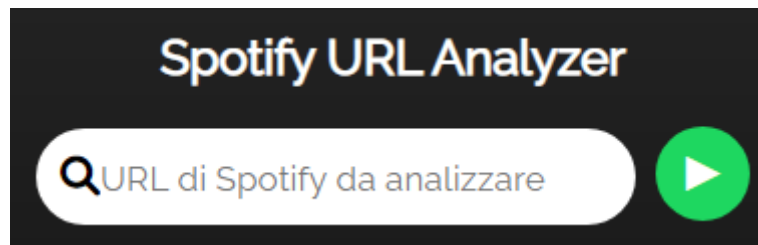


# Spotify URL Analyzer

Saverio Cannone – Mat. 700567

Repository - [https://github.com/Xav147/ICon\\_2021\\_Cannone](https://github.com/Xav147/ICon_2021_Cannone)



## Dataset

Il dataset utilizzato è disponibile al seguente link:

<https://www.kaggle.com/datasets/yamaerenay/spotify-dataset-19212020-600k-tracks>

Quest'ultimo contiene più di 600.000 tracce musicali di Spotify e i dati di oltre 1 milione di artisti.

Per poter utilizzare al meglio il dataset è necessario svolgere alcune operazioni per pulire i dati.

## Data Preparation

Alcune feature del dataset non sono prontamente utilizzabili. Per esempio, **release\_date** è una stringa che può essere trasformata nella decade della canzone.

Inoltre, il dataset non offre informazioni sul genere di una canzone; tuttavia, sono associati dei generi per gli artisti. Questo permette di ricavare il genere di una canzone in base all'artista.

## Funzionalità

Le funzionalità principali del programma sono:


- **Raccomandazione** di un insieme di canzoni in base all'analisi della playlist inserita dall'utente
- **Classificazione** del genere di una canzone inserita dall'utente
- **Raccomandazione** di canzoni simili a quella inserita dall'utente

L'approccio utilizzato per la **raccomandazione** di una playlist è il **TF-IDF (Term Frequency – Inverse Document Frequency)**.

Con il **TF-IDF** è possibile dare più o meno peso ai metadata sul genere a seconda di quanto spesso determinati generi compaiono nel dataset.

Una volta generato un set di feature basandosi sul **TF-IDF**, è possibile rappresentare ogni canzone in una playlist come un vettore di features. È possibile dunque utilizzare la **cosine similarity** per individuare canzoni che sono simili alla playlist analizzata.

# Analisi di una playlist



PLAYLIST

## The most famous songs

All those songs you've heard at least once in your life


Gianlucadibari • 4.568 Mi piace • 257 brani, 15 ore 44 min.





Q Ordinamento personalizzato ▾


Una volta fornito l'URL della playlist che si vuole analizzare il programma genera un'altra playlist con canzoni che sono simili al vettore che rappresenta la playlist.


### Playlist analizzata


**Payphone**  
Maroon 5

**Play Hard (feat. Ne-Yo & Akon) - New Edit**  
David Guetta


**7 Years**  
Lukas Graham


**Jubel - Radio Edit**  
Klingande


**How to Save a Life**  
The Fray


**Stressed Out**


### Canzoni consigliate


**Bohemian Rhapsody - Remastered 2011**  
Queen

**Hey Jude - Remastered 2015**  
The Beatles

**Piano Man**  
Billy Joel

**Heaven**  
Bryan Adams

**Imagine - Remastered 2010**  
John Lennon

**Your Song**

# Canzone analizzata



## Beat It

Michael Jackson

## Audio features

Danceability: 0.779

Energy: 0.867

Key: 3

Loudness: -3.704

Mode: 0

Speechiness: 0.0457

Acousticness: 0.0491

Instrumentalness: 7.98e-06

Liveness: 0.197

Valence: 0.915

Tempo: 138.858

## Analisi di una canzone

Per la classificazione del genere di una canzone è stato necessario semplificare il dataset, in quanto contenente molti generi con pochi esempi oppure esempi senza alcun genere. Raccogliendo solo i 20 generi più rappresentati, il totale delle tracce si attesta sui 350.000. I modelli di

### **apprendimento supervisionato**

usati per la classificazione sono:

- **Logistic Regression**

Modello di classificazione che punta a minimizzare **log-loss** per determinare i pesi di una **funzione lineare**

- **Neural Networks**

È stata utilizzata una rete neurale di tipo **MLPC** (Multi-layer Perceptron Classifier), che va a minimizzare la **log-loss**.

- **Random Forest**

Algoritmo di **ensemble learning** che genera più learner (alberi) per effettuare predizioni, unendole alla fine per ottenere il classificatore.

Il programma mostra le **percentuali** di probabilità del genere, ordinate in modo decrescente.

Oltre a classificare il genere della canzone, il programma fornisce dei suggerimenti in base alla canzone analizzata.

Per generare i suggerimenti viene utilizzato **l'apprendimento non supervisionato**, in particolare vengono consigliate canzoni che appartengono allo stesso cluster della canzone analizzata. I cluster vengono generati utilizzando **k-means clustering**, con  $k = 1000$ .

## Canzoni consigliate

Hear it grow - Underground Sound Of Lisbon  
False Alarm - The Weeknd  
The Shortest Message (2015) - Todd Glass

## Analisi del genere

pop: 20%  
hoerspiel: 18%  
rock: 16%  
easy\_listening: 11%  
classic: 9%  
jazz: 8%  
dance: 5%  
soul: 4%  
metal: 3%  
latin\_alternative: 2%  
hip\_hop: 1%  
orchestral\_performance: 1%  
latin: 1%  
disco: 1%  
rap: 0%  
folk: 0%  
country: 0%  
punk: 0%  
filmi: 0%  
progressive\_house: 0%