

Liking Songs

Un progetto di Fondamenti di Intelligenza Artificiale

Autore: Giovanni Citro

Indice

1	Introduzione	2
2	Definizione del Problema	3
3	3.1 Performance	4 4 4 4 4
4	4.2 Analisi e scrematura del dataset	5 5 5 6
5	Modello di Classificazione	7

1 Introduzione

Il sistema proposto è un modello di classificazione che si basa sull'interazione attiva degli utenti per predire le loro preferenze musicali. Gli utenti inseriscono le canzoni preferite, assegnando feedback positivi (1) o negativi (0). Questi dati addestrano un modello avanzato che apprende in modo personalizzato le preferenze di ciascun utente.

2 Definizione del Problema

Il sistema implementato si basa su un sofisticato modello di classificazione che permette agli utenti di inserire le proprie canzoni preferite all'interno di una lista apposita. La fase successiva coinvolge l'utente nel processo decisionale, chiedendogli di assegnare un feedback positivo (1) o negativo (0) per ciascuna canzone, indicando se la stessa gli piaccia o meno. Questo feedback utente diventa fondamentale per addestrare il modello, che utilizza i dati raccolti per apprendere e comprendere le preferenze musicali specifiche di ciascun utente.

Il sistema, sfruttando le informazioni fornite dall'utente durante la fase di classificazione delle canzoni preferite, è in grado di creare un modello personalizzato di predizione delle preferenze musicali. In altre parole, il sistema apprende i tratti distintivi delle canzoni gradite e non gradite da parte dell'utente in questione, consentendo predizioni future sul gradimento di nuove canzoni che non sono state precedentemente valutate.

3 Specifica dell'ambiente

3.1 Performance

La misura di performance dell'agente è valutata in base alla sua abilità di avvicinarsi il più possibile a una situazione ideale, in cui agli utenti vengono mostrati esattamente i valori che indicano se una canzone piace (1) o non piace (0). Questa valutazione è effettuata utilizzando il dataset come riferimento, con l'aggiunta della colonna "Liking" che rappresenta il valore ideale di gradimento associato a ciascuna canzone.

3.2 Environment

L'ambiente in cui opera l'agente è costituito dall'insieme di canzoni con i relativi dati di gradimento. Le canzoni presenti nell'ambiente sono caratterizzate dai feedback degli utenti, rappresentati attraverso i valori binari di gradimento (piace o non piace), e sono descritte nel dataset utilizzato per l'addestramento e la valutazione del modello.

3.3 Actuators

Gli attuatori dell'agente consistono nella lista di canzoni con un parametro che ci dice quale canzone può piacere e quale no

3.4 Sensors

I sensori rilevano il feedback degli utenti, espressi attraverso i valori binari di gradimento (piace o non piace), per tutte le canzoni presenti nell'ambiente. Questi dati sensoriali sono fondamentali per l'addestramento e la valutazione del modello di classificazione, permettendo all'agente di apprendere e migliorare la sua capacità di predire le preferenze musicali degli utenti.

4 Raccolta, Analisi e Scrematura del dataset

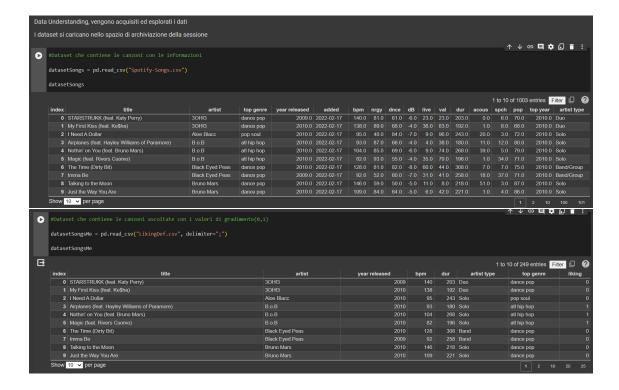
4.1 Scelta del dataset

Per i dataset del modello di classificazione ho usato la strada più semplice, ovvero quella di cercare un dataset con le canzoni di spotify e i suoi attributi. Infine ho creato un dataset inserendo la colonna che mi serviva per far apprendere al modello i valori di gradimento da predire

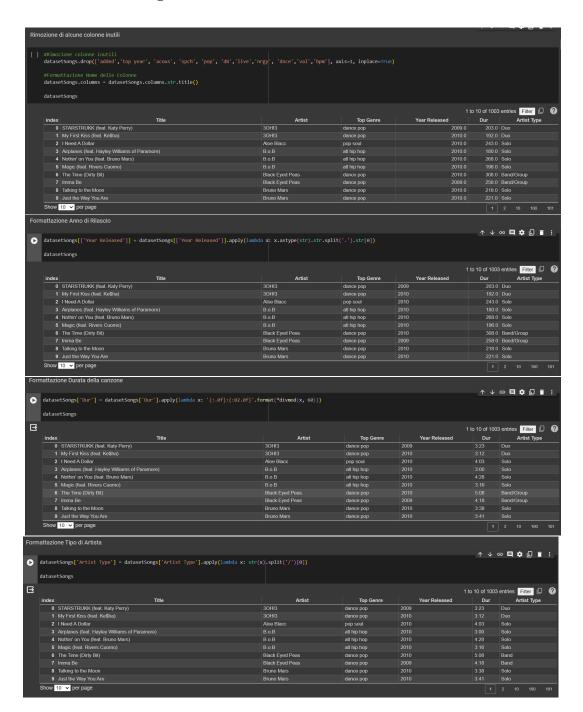
4.2 Analisi e scrematura del dataset

Guardando bene il dataset ho visto che alcune colonne erano inutili al mio intento e quindi sono andato ad eliminarle(added, top year, acous, spch, pop, dB, live, nrgy, dnce, bpm), poi ho formattato alcuni valori all'interno di alcune colonne come: Anno di Rilascio, Durata e Tipo di Artista.Il progetto completo fatto su Colab è presente sul mio GitHub.

4.3 Data Understanding



4.4 Data Cleaning



5 Modello di Classificazione

Questo modello utilizza un Decision TreeClassifier per classificare se una persona gradisce o meno una canzone basandosi sulle caratteristiche fornite nel Data Frame. La preparazione dei dati, la codifica delle variabili categoriche, la divisione dei dati e la valutazione delle prestazioni sono passaggi cruciali in questo processo. In oltre uso il One Hot Encoding che è una codifica per convertire le variabili categoriche in array di soli 0 ed 1 poichè alcuni modelli potrebbero non accettare valori diversi da quelli numerici, in quanto le stringhe per loro non hanno significato.

Come ultimo passaggio verifichiamo i risultati che ha ottenuto grazie ai dati che avevamo messo da parte con il test set. Come accuratezza abbiamo una percentuale che va dallo 0.76 all'1.00 e come precisione dallo 0.75 al 1.00