MoodUp: songs advice based on your mood

Giorgia Gossi, Aurora Musitelli

Statistica spaziale e ambientale - a.a. 2021/2022

1 ABSTRACT

Il seguente report, relativo al progetto realizzato, propone l'analisi riguardante i testi delle canzoni di vari generi musicali pubblicate dal 1950 al 2019 prendendo in considerazione il dataset https://www.kaggle.com/datasets/saurabhshahane/music-dataset-1950-to-2019. All'interno del dataset vi è la topic analysis realizzata mantenendo 7 livelli della variabile target multiclasse, ed è stata creata attraverso la tecnica LDA. La topic analysis è stata utile per la classificazione delle canzoni e, una volta vettorizzati i testi delle canzoni attraverso TF-IDF, si è utilizzato il modello migliore random forest per trovare le parole più importanti dei testi delle canzoni ed estrarre le canzoni che sono state correttamente classificate per creare un modello logistico surrogato come XAI. Come obiettivo finale si è creato il chatbot, **MoodUp,** in grado di consigliare agli utenti più di una canzone in base allo stato d'animo/al mood che gli utenti esprimono attraverso l'utilizzo di 3 parole.

2 INTRODUZIONE PROGETTO

La musica ha sempre fatto parte dell'esperienza umana e soprattutto, al giorno d'oggi, i servizi di streaming musicali sono sempre più diffusi e cercano di specializzarsi sempre di più sulle richieste degli utenti creando delle playlist specializzate in base ai propri ascolti giornalieri. Molti dei comportamenti sociali si esprimono attraverso le trasformazioni musicali che si intrecciano con il tempo e proprio per questo è utile analizzare le interazioni tra il processo di evoluzione musicale insieme ai progressi tecnologici che possono permettere di creare delle innovazioni. Nel seguente progetto, quindi, si è pensato di rispondere ad una necessità degli utenti, avere un sistema che consigli delle canzoni in base al proprio umore/stato d'animo.

3 MATERIALI UTILIZZATI

Il punto di partenza è stato il paper "Temporal Analysis and Visualisation of Music" di Gomes de Moura et al. dove è stato creato il dataset da noi utilizzato.

I dati utilizzati in questo paper provengono dal pacchetto spotipy di Python e dall'API di lyrics genius per scaricare i testi delle canzoni. Come dati di riferimento sono state considerate 82452 canzoni distribuite su 7 generi musicali pubblicate tra il 1950 e il 2019. Per ciascuna di queste canzoni è stato scaricato il testo e delle features audio fornite da spotify; queste le feature audio selezionate:

- acousticness rappresenta la presenza di strumenti in acustico
- danceability punteggio di quanto la canzone è adatta a ballare
- loudness volume medio della canzone in decibel (dB)
- instrumentalness descrive se la canzone contiene strumenti musicali
- valence indica l'umore della canzone
- energy identifica l'energia della canzone

E' stata svolta una pulizia del testo, sono stati rimossi i testi non in inglese, sono stati rimossi elementi non necessari, come le onomatopee e il nome dell'artista che canta la strofa, stopwords ed è stata utilizzata la lemmatization.

Sui testi puliti è stata svolta la topic analysis utilizzando la tecnica Latent Dirichlet Allocation (LDA); questo modello si fonda sull'idea che la distribuzione di probabilità dei documenti è basata su dei topic latenti, la cui distribuzione è caratterizzata da parole. Il numero ottimale di topic è risultato essere 19, quali: dating, night/time, violence, world/life, shake the audience, family/gospel, romantic, communication/thinking, obscene, sound, movement/places, light/visual perceptions, family/spiritual, like/girls, sadness, feelings. Nel dataset fornito dal paper sono presenti 28372 canzoni e sono stati mantenuti soltanto gli 8 topic sadness, violence, world/life, obscene, music, night/time, romantic e feelings.

4 LAVORO SVOLTO

Dopo aver importato il dataset abbiamo svolto alcune analisi preliminari per controllare la corretta importazione dei dati e che non fossero presenti dei dati mancanti. Inoltre, dopo aver visionato alcune canzoni per ciascun livello del target, è stato deciso di unire i due livelli "world/life" e "feelings" in un unico livello chiamato "life", data la similarità dei contenuti; in questo modo la variabile target multiclasse presenta sette livelli; questa la sua distribuzione:

```
sadness      6096
life         6032
violence     5710
obscene      4882
music        2303
night/time   1825
romantic     1524
```

Si è poi proceduto con la pulizia dei testi, che erano comunque già stati pre-processati dagli autori del paper, applicando la lemmatizzazione, aggiungendo bigrammi ed n-grammi e rimuovendo le stopwords, a cui abbiamo aggiunto alcune parole come onomatopee o descrizioni delle canzoni. Successivamente sono stati rimossi alcuni bigrammi che erano formati dalla stessa parola (es. "life-life", "sing-sing") e quattro canzoni il cui testo era formato da meno di cinque parole. Prima di stimare il modello si è applicata la vettorizzazione dei testi; sono stati provati tre diversi metodi: bag of words, tf-idf e bert. Per la stima del modello abbiamo provato sia con una random forest che con una rete neurale, in combinazione con i diversi metodi di vettorizzazione e cambiando i parametri dei modelli. Utilizzando la cross validation a tre fold, il modello migliore è risultato essere una vettorizzazione con tf-idf, che produce vettori di lunghezza 1000, in combinazione con una random forest formata da 100 alberi, che utilizza il criterio dell'impurità di Gini, senza una profondità massima e con minimo due osservazioni per la divisione del nodo. Il metodo tf-idf trasforma ciascun documento in un vettore, dove ogni elemento i del vettore j è dato da:

$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$ dove $tf_{i,j}$ rappresenta il numero di occorrenze della parola i nel documento j, df_i il numero di documenti che contengono la parola i e N il numero di documenti j. Il modello random forest si basa sull'estrazione di più campioni bootstrap su cui vengono stimati degli alberi che sfruttano soltanto un subset dei predittori disponibili; la classificazione finale è data dal majority vote.

Queste le performance del modello, calcolate sul dataset di test:

	precision	recall	f1-score	support
life	0.54	0.54	0.54	1809
music	0.50	0.55	0.53	691
night/time	0.38	0.38	0.38	547
obscene	0.72	0.77	0.74	1465
romantic	0.45	0.40	0.42	457
sadness	0.56	0.55	0.56	1829
violence	0.61	0.59	0.60	1713
accuracy			0.57	8511
macro avg	0.54	0.54	0.54	8511
weighted avg	0.57	0.57	0.57	8511

Figura 1

Per verificare il corretto funzionamento del modello e capire quali fossero le parole più importanti per la classificazione delle canzoni in ciascun livello del target è stato utilizzato un modello logistico surrogato; abbiamo quindi applicato un modello logistico alle predictions fatte dalla random forest, utilizzando soltanto le osservazioni del dataset di training correttamente classificate dalla random forest.

Le metriche del modello:

	precision	recall	f1-score	support
life	0.62	0.70	0.66	4170
music	0.74	0.58	0.65	1591
night/time	0.62	0.35	0.45	1274
obscene	0.82	0.84	0.83	3409
romantic	0.68	0.41	0.51	1048
sadness	0.65	0.71	0.68	4203
violence	0.71	0.77	0.74	3977
accuracy			0.69	19672
macro avg	0.69	0.62	0.64	19672
weighted avg	0.69	0.69	0.69	19672

Figura 2

Per ciascuno dei 7 livelli del target sono stati ottenuti i coefficienti del modello logistico più alti in valore assoluto e le parole ad essi abbinate. Abbiamo deciso di utilizzare le 85 parole più importanti per ciascun livello dopo aver visionato l'andamento dei coefficienti attraverso dei plot e cercando di mantenere soltanto quelli che si posizionavano prima del gomito.

Qui, ad esempio, il plot per il livello “night/time”:

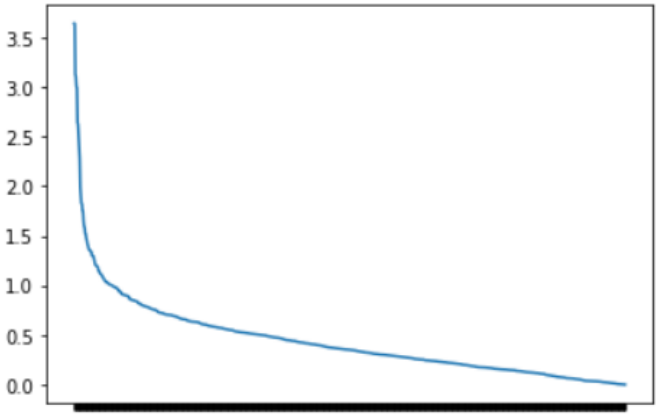


Figura 3

E’ stata analizzata la relazione tra i diversi topic, verificando quante parole avessero in comune tra le 85 parole precedentemente selezionate.

	life	music	night_time	obscene	romantic	sadness	violence
life	85	10	11	6	12	3	9
music	10	85	10	13	5	10	9
night_time	11	10	85	7	2	14	10
obscene	6	13	7	85	6	7	3
romantic	12	5	2	6	85	10	8
sadness	3	10	14	7	10	85	8
violence	9	9	10	3	8	8	85

Figura 4

I topic che presentano più parole in comune sono: “night/time” e “sadness” mentre i topic che presentano meno parole in comune sono: “night/time” e “romantic”.

Nel dataset di partenza erano presenti anche le features audio fornite da Spotify, quali 'danceability', 'loudness', 'acousticness', 'instrumentalness', 'valence', 'energy'; si tratta di punteggi compresi tra 0 e 1 e che abbiamo suddiviso in quattro classi ([0, 0.25], (0.25, 0.5], (0.5, 0.75], (0.75, 1]). E’ stata analizzata l’associazione tra queste e i topic dei testi; innanzitutto abbiamo calcolato l’indice Chi Quadro normalizzato di Pearson con questi risultati:

```
Chi quadro danceability:          0.22608164276368710
Chi quadro loudness:              0.10777018410175673
Chi quadro acousticness:          0.21937621015073594
```

```
Chi quadro instrumentalness:     0.03149316071558962
Chi quadro valence:              0.08285476954140855
Chi quadro energy:               0.22505729364247612
```

Nel caso delle associazioni più alte è stata svolta un'analisi delle corrispondenze per verificare quali livelli delle variabili fossero maggiormente associati tra loro. In questi grafici sono rappresentate le prime due componenti.

Danceability

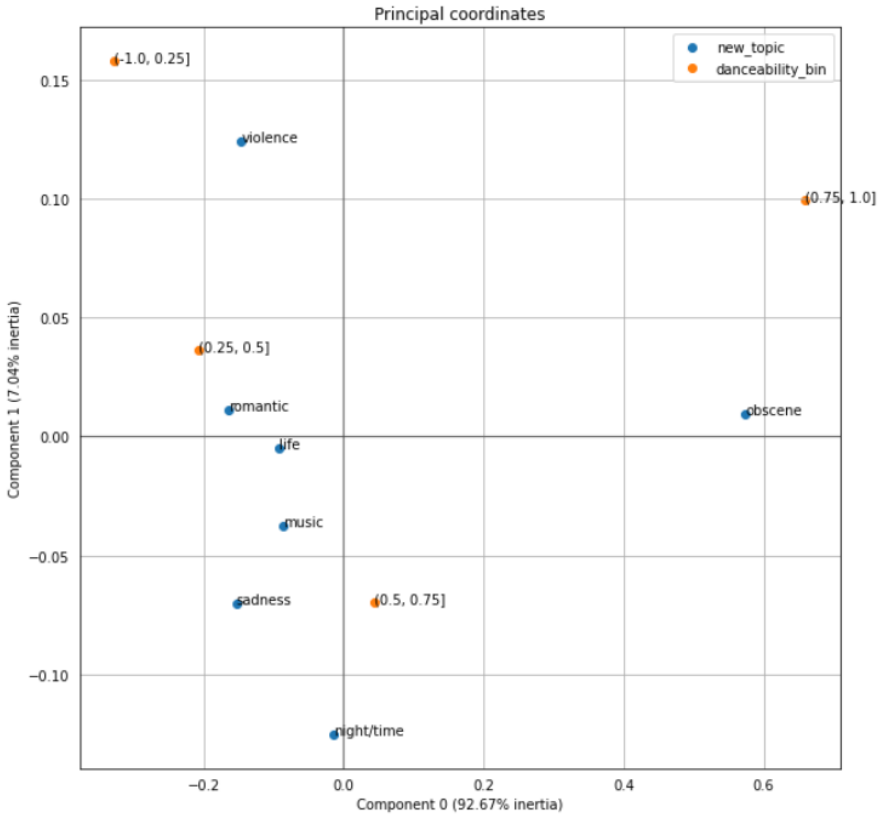
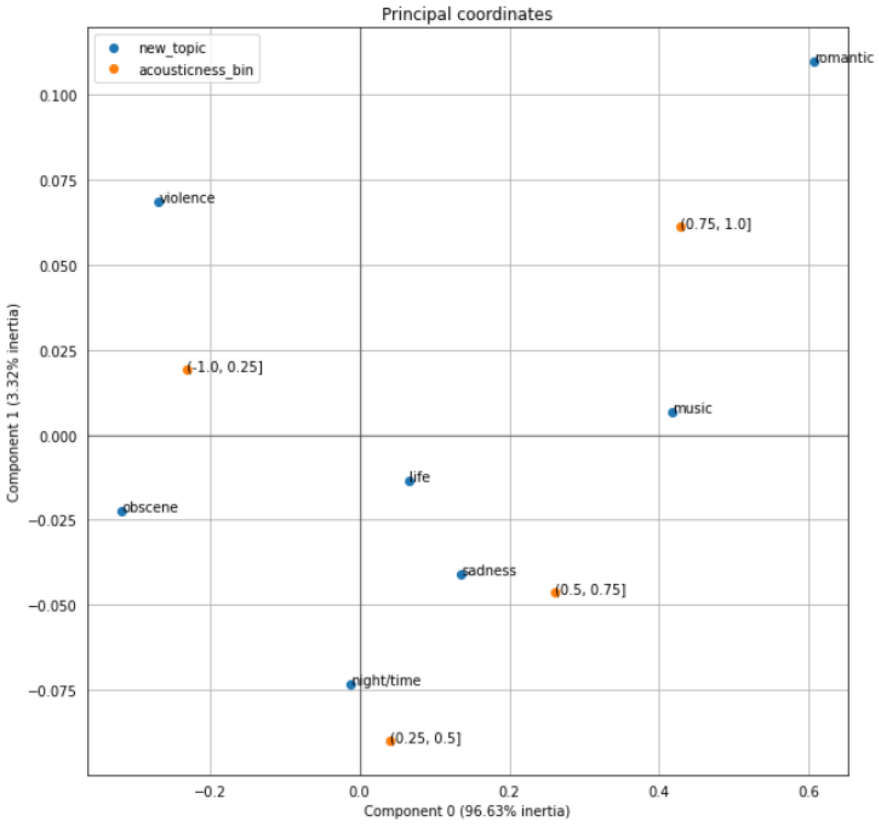


Figura 5

Possiamo notare che le due componenti spiegano più del 99% dell'associazione tra le due variabili. Il topic violence ed un basso punteggio di danceability sembrerebbero essere tra loro associati; anche night/time ed un punteggio medio alto di danceability sembrerebbero essere piuttosto associati.

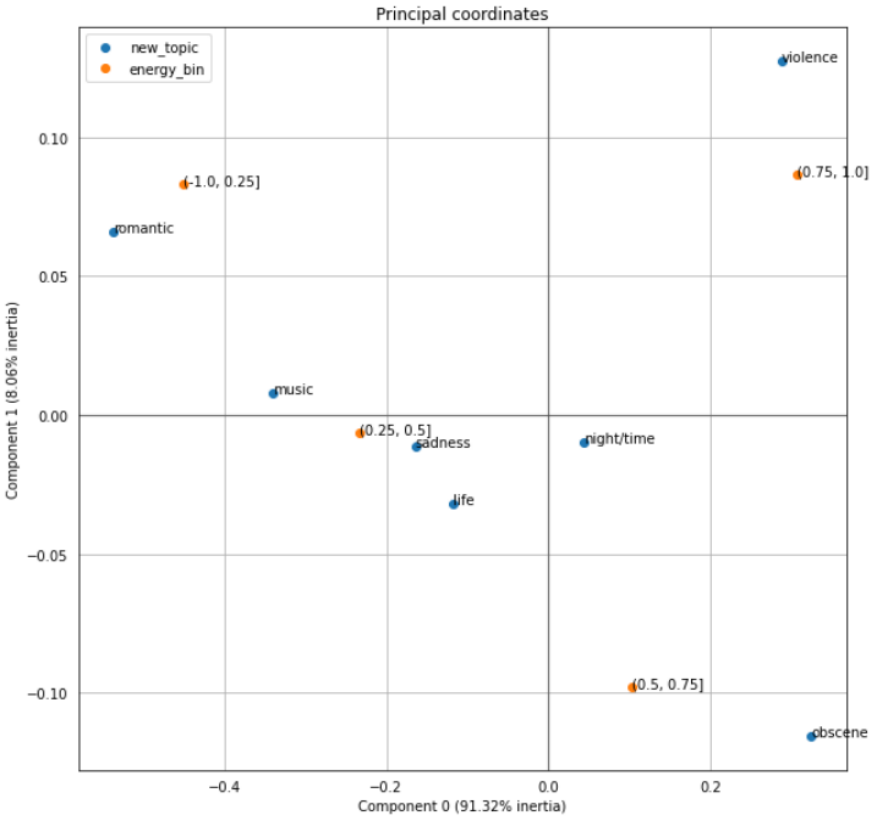
Acousticness



Anche in questo caso le due componenti spiegano più del 99% dell'associazione. Qui il topic night/time sembrerebbe essere associato con un punteggio di acousticness compreso tra 0.25 e 0.75 mentre il topic romantic con un punteggio tra 0.5 e 0.75.

Figura 6

Energy



Le due componenti riescono nuovamente a spiegare la quasi totalità dell'associazione. In questo caso una bassa energia sembra essere associata al topic romantic mentre un'alta energia al topic violence; inoltre il topic obscene sembra essere in relazione con un punteggio medio alto di energia.

Figura 7

5 RISULTATI

I risultati più rilevanti, delle analisi che si sono svolte, sono i grafici realizzati relativi alla prevalenza dei diversi 7 topic rispetto alle canzoni del dataset. In particolare, sono state estratte le probabilità previste dei 7 livelli per le canzoni tramite l'attributo "predict_proba", queste probabilità previste derivano dalla classificazione del modello migliore, random forest, del dataset di test. Si è quindi deciso di prendere in analisi 3 canzoni per creare i rispettivi grafici che indichino il mood relativo ai 7 topic. In particolare le canzoni scelte sono state:

1. This is how we do, Katy Perry - classificata correttamente "obscene"
2. Mr blue sky, Electric light orchestra - classificata correttamente "music"
3. Happy together, The turtles - classificata come "sadness" ma in realtà "life"

Utilizzando la libreria matplotlib e inserendo le caratteristiche specifiche del grafico, prendendo in considerazione come values le probabilità previste del modello del dataset di test si ottengono i seguenti grafici:

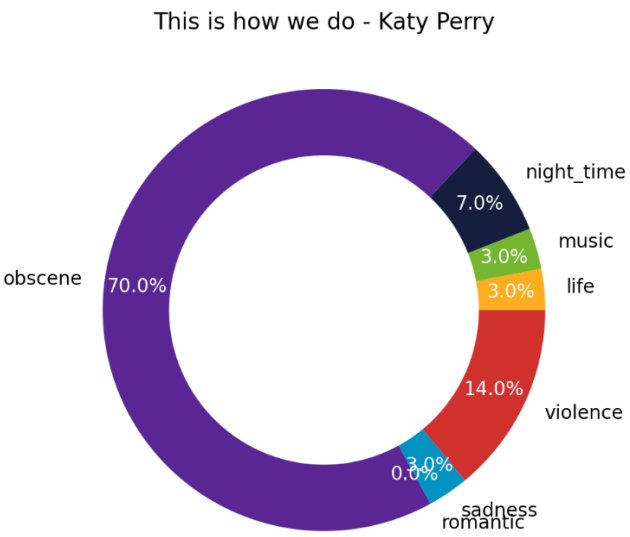


Figura 8

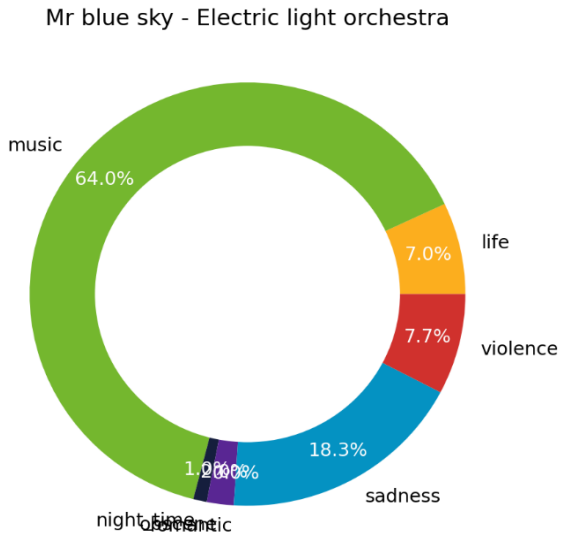


Figura 9

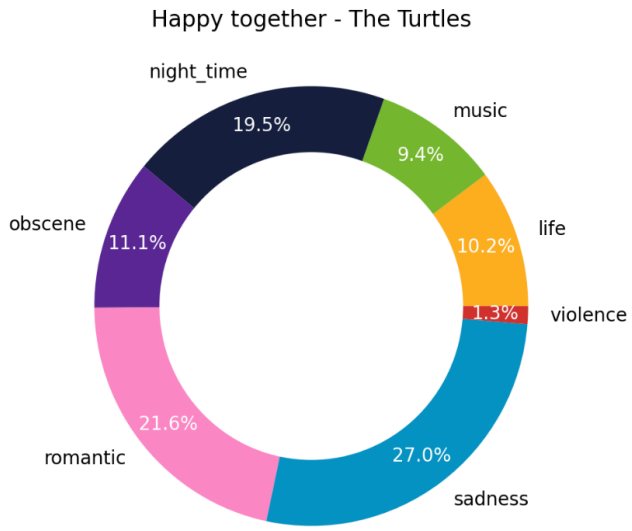


Figura 10

Osservando i tre grafici si può notare la prevalenza dei vari topic nelle canzoni selezionate, infatti per le prime due canzoni correttamente classificate prevale il topic di riferimento. Mentre per l'ultima canzone che non è stata classificata correttamente si può osservare una prevalenza netta del topic "sadness" anche se nel dataset completo la canzone presenta un topic rilevante che è "life".

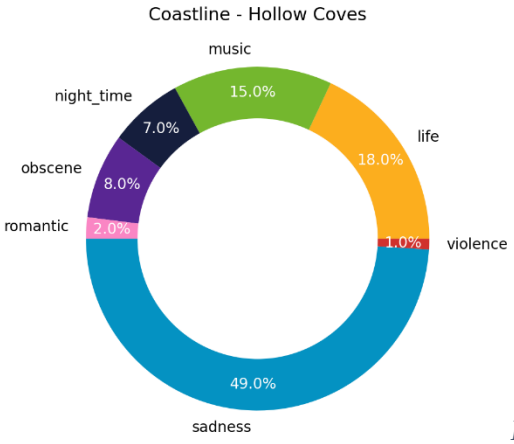


Figura 11

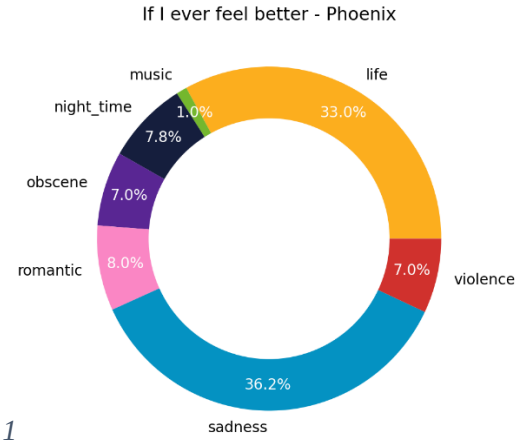


Figura 12

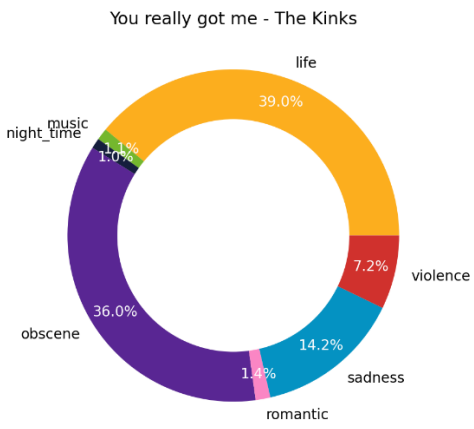


Figura 13

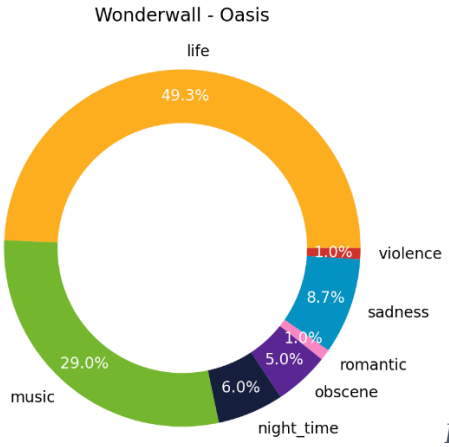


Figura 14

Abbiamo costruito un piccolo dataset di score composto da 4 canzoni a cui abbiamo applicato il modello ottenendo le probabilità previste dei topic; queste sono rappresentate nei grafici dalla *figura 11* alla *figura 14*. Nel caso delle canzoni "Coastline" e "you really got me" la classificazione non sembra essere del tutto corretta, nel caso della prima canzone la componente life dovrebbe predominare mentre nella seconda canzone la componente obscene dovrebbe essere la maggiore e la componente romantic dovrebbe avere un peso più alto.

6 CONCLUSIONE: realizzazione chatbot MoodUp

Come obiettivo finale del progetto si è deciso di realizzare il chatbot di telegram chiamato **MoodUp** che ha come funzione principale quella di rispondere alle richieste degli utenti consigliando le canzoni rispetto al mood/stato d'animo delle parole inserite dall'utente.

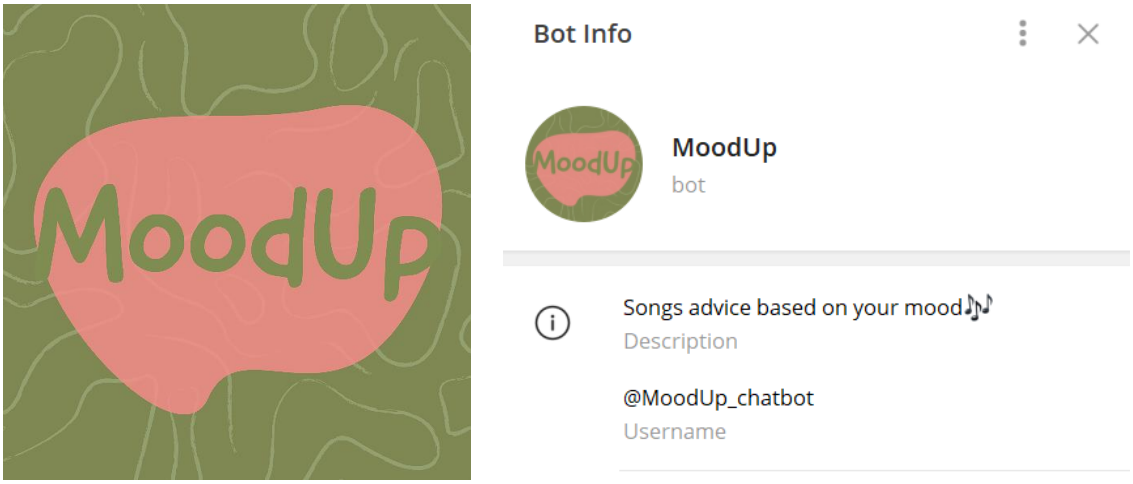


Figura 11

Il nome del chatbot, il logo e la descrizione sono stati pensati e realizzati pensando ad un target giovanile che ascolta musica di vario genere e diversificato negli anni. Attraverso il logo si vuole restituire un ambiente friendly di good vibes, le onde rappresentano le emozioni/stato d'animo poichè questo chatbot viene utilizzato nel momento in cui gli utenti si sentono a corto di idee sul tipo di musica da ascoltare e vogliono avere dei consigli musicali in base al loro umore. Il nome del chatbot si basa sull'idea di "alzare il proprio mood/stato d'animo" e i colori vogliono richiamare la veste vintage e moderna delle canzoni rispetto all'arco temporale che è stato preso come riferimento. La descrizione del chatbot rispecchia perfettamente il focus di questo strumento, un modo semplice e veloce per poter avere un consiglio musicale, da condividere con amici e persone strette per poter scambiarsi a vicenda nuova musica.

Per capire il funzionamento del chatbot si fornisce una descrizione del lavoro svolto; Una volta che l'utente inserisce nella chat "**/start**" il chatbot restituisce il messaggio della *figura 12*

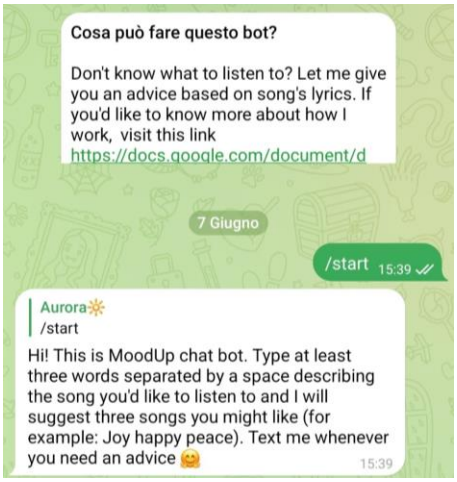


Figura 12

Successivamente l'utente deve inserire 3 parole relative alle canzoni che vorrebbe ascoltare in base al suo stato d'animo. Queste parole inserite dall'utente vengono considerate come il testo di una canzone e quindi vettorizzate e classificate dal modello random forest che è stato utilizzato. Le canzoni che vengono consigliate sono quelle che hanno un vettore di probabilità previste più vicino a quello della "canzone" inserita dall'utente. Inoltre per una maggiore efficienza del chatbot si è lavorato per singola parola utilizzando gli embeddings. In particolare è stato scelto FastText invece che Word2vec in quanto permette di processare anche parole non presenti nel suo vocabolario; questo modello è stato allenato su tutti i testi delle canzoni presenti nel dataset già puliti e preprocessati. FastText si basa comunque sugli algoritmi di Word2vec, sfrutta quindi una rete neurale con metodo di attivazione hierarchical softmax; nel nostro caso è stato utilizzato cbow (continuous bag of words) dove si utilizza il contesto per predire la parola centrale. Si è deciso di utilizzare cbow invece di skipgram data l'ampiezza dei nostri dati e anche perché non è necessario rappresentare le parole più rare dato il nostro utilizzo di word embeddings. Utilizzando gli embeddings si considera anche l'aspetto lessicale e semantico delle parole e il testo inserito è così ampliato per poter ricevere un consiglio più pertinente rispetto allo stato d'animo che l'utente vuole esprimere. La scelta di utilizzare gli embeddings è significativa, poiché se l'utente inserisce delle parole che non sono presenti nel vocabolario delle parole delle canzoni, il vettorizzatore non saprebbe come gestirle e restituirebbe consigli delle canzoni non inerenti allo stato d'animo espresso.

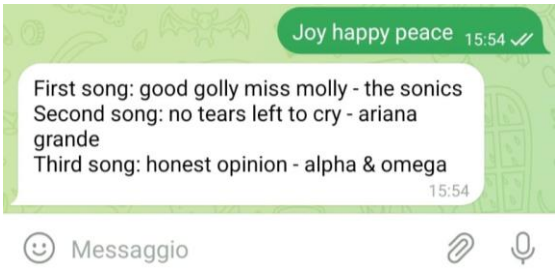


Figura 13

7 CRITICITA' e PROSPETTIVE DI MIGLIORAMENTO

Le **criticità** maggiori sono state riscontrate:
Nella selezione del modello e del metodo di vettorizzazione migliori perché sono state svolte diverse prove e inoltre abbiamo dovuto effettuare una scelta riguardo l'utilizzo della versione della libreria Gensim, in quanto le versioni 3 e 4 fornivano risultati diversi in particolare con la funzione "most_similar". Si è deciso di optare per la versione più recente poiché permette di elaborare più parole; ad esempio la parola "you", rimossa durante la pulizia del testo in quanto stopwords, genera un errore utilizzando Gensim 3 mentre è processata dalla versione 4. Nella scrittura del codice per il chatbot abbiamo comunque inserito un'eccezione nel caso in cui tale errore si verifichi anche con la versione più nuova in modo che il chatbot possa comunque funzionare.

Le **prospettive di miglioramento** del chatbot MoodUp sicuramente possono essere quelle di ampliare il dataset delle canzoni in modo tale da avere maggiori consigli musicali che può restituire il chatbot, anche per ampliare il target di riferimento per fare in modo che più persone possibili potrebbero utilizzarlo.

8 APPENDICE

I 3 notebook relativi alle analisi e il codice relativo al chatbot.