ERAS: Emotion Recognition of Audio Signals

Cos'è ERAS?

Diamogli una semplice definizione!

ERAS è un Al capace di predire

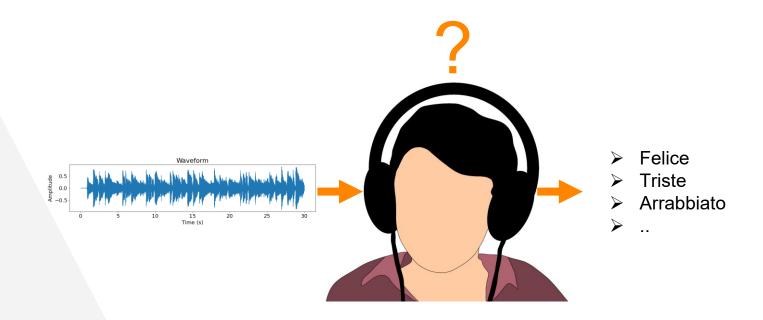
le emozioni che la musica ci

fa provare!

2. Problematica

Qual è la difficoltà in ciò?

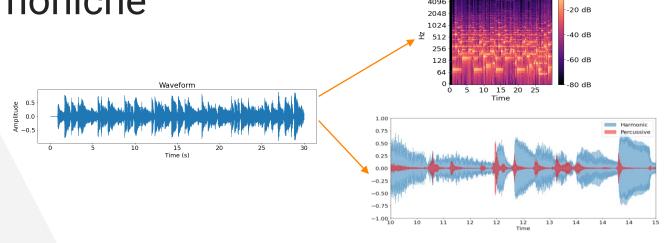
Problematica: percezione umana



Problematica: componenti segnale audio

Trasformata di Fourier

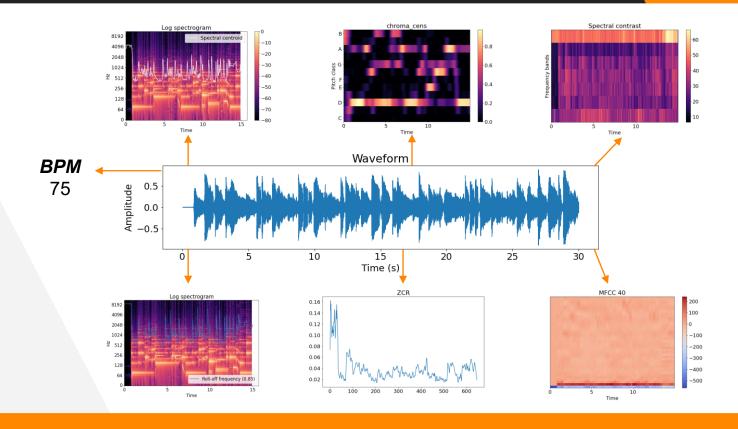
Componenti percussive & armoniche



Spectrogram

+0 dB

Problematica: features audio



3. Dataset

Creazione del dataset

Dataset: Youtube Audio Library

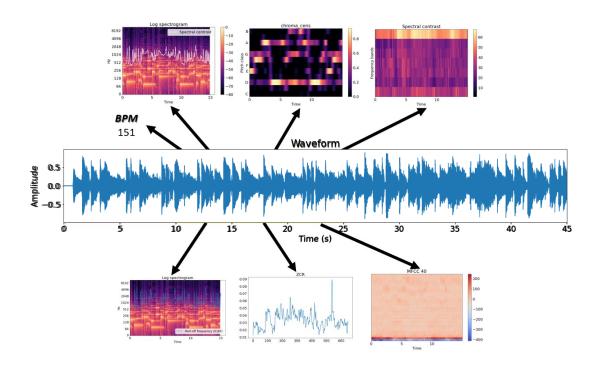
Vantaggi:

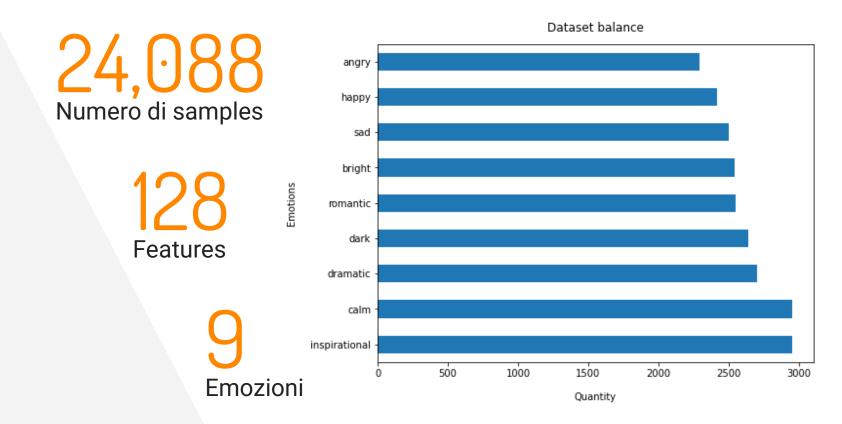
- Free royalty music
 - Divise per emozioni
 - Grande quantità



Dataset: Estrazione features

Finestra: 15s Salto: 10s





4. Approccio

Step di sviluppo usati

Approccio: tabella riassuntiva

Attempt	Features	N° of features	Audio's seconds	F1 macro scores	Notes
1st	chroma stft, rmse, spectral centroid, spectral bandwidth, rolloff, zcr, MFCCs 13	26	30s	SVC: 21% KNN: 19%	Lack of features and samples
2nd	Spectral Chroma, MFCCs 13, Spectral Centroid, Spectral Contrast, Spectral RollOff, ZCR, BPM	74	Whole track duration	SVC: 26% KNN: 21%	Lack of samples
3rd	Spectral Chroma, Mfccs 13, Spectral Centroid, Spectral Contrast, Spectral RollOff, ZCR, BPM	74	Window size: 15s Hop size: 10s	SVC: 77% KNN: 90%	Check the dataset & model validity
Val.	Spectral Chroma, Mfccs 13, Spectral Centroid, Spectral Contrast, Spectral RollOff, ZCR, BPM	74	Window size: 15s Hop size: 10s	KNN k-cross validation: 89%	/
Final	Spectral Chroma, Mfccs 40, Spectral Centroid, Spectral Contrast, Spectral RollOff, ZCR, BPM	128	Window size: 15s Hop size: 10s	KNN k-cross validation: 92%	/

Table 4.1: Summary table

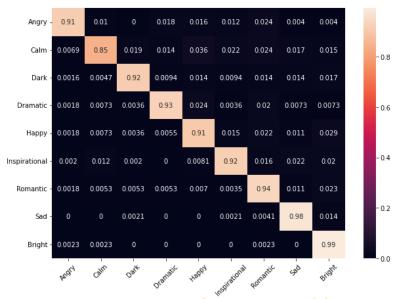
Approccio: matrici di confusione

Primo risultato



F1 Macro Score: 19%

Risultato finale



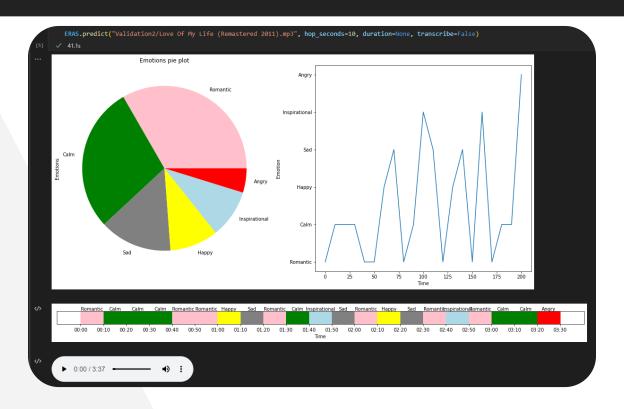
F1 Macro Score: 92%

Approccio: descrizione architettura

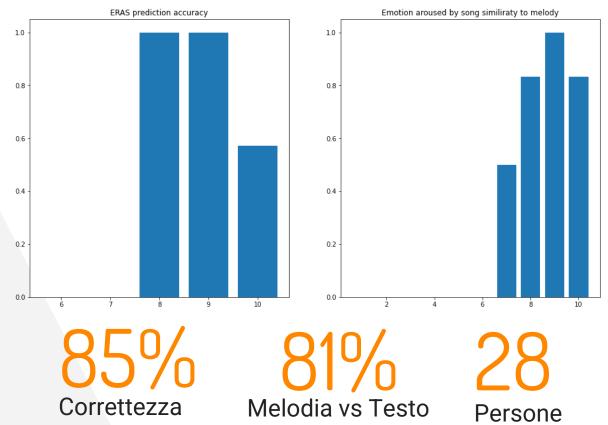
ERAS model: Model scaler: Scaler encoder: Encoder dataset: String extractor: Extractor + train(test_size:float, showScore: boolean): ERAS + predict(track:String, offset:float, duration:float, sr:int,window_seconds:int, hop seconds:int,transcribe:boolean):void + voiceEmotion(sr:int, seconds:int):void + setDataset(dataset:String):void dataPreProcessing(test_size:float) modelScore(X test:numpyArray<float>, y test:numpyArray<float>):void __transcribe(y:list<float>):void - __printResult(prediction:list<String>,window_seconds:int,hop_seconds:int):void

Extractor - __emotions: list<String> - __mfccs_n: int - __header: list<String> + featuresExtractor(song_name:string, y:list<float>, sr:int):list<String> + featuresExtractorWindowed(track:String, sr:int, window_seconds:int,hop_seconds:int):numpyArray<float> + datasetMaker(path_tracks:String, filename:String, encoder, sr:int, window_seconds:int, hop_seconds:int, print_metadata:boolean):void - __headerMaker():list<String> - __csvMaker(filename:String, data:list<String>, perms:char):void

ERAS: esempio di utilizzo







Grazie mille per l'attenzione!