Lyrics: authorship and emotion analysis

Anca Sotir, Florin Manghiuc, Irina Chitu

January 17, 2022

Abstract

Each artist has his own style and personality which greatly impacts his work. The same message can be expressed completely different depending on who sings it. This project explores several machine learning methods to solve the task of identifying the author of a given text (attribution of authorship). **[[insert some names + some results]]**

Using web scraping we fetched data from a lyrics site. For a better understanding of our task, we paid close attention to our custom dataset. We analysed the emotions and the similarity between different songs of the same artist but also between various artists.

Key words: scraping, authorship, lyrics, emotion, similarity

1 Introduction - yes

2 Dataset

Our dataset was made by scraping the azlyrics site [AZL]. While fetching the data using BeautifulSoup [Bea], we encountered several issues such as missing lyrics or songs wrongly assigned. Furthermore, after too many requests, the site was blocking our access so we had to add some delays between our requests and also use an agent. The data retrieval was followed by a small clean-up: songs in another language than English were dropped. The filtering was done using spacy_langdetect.LanguageDetector. A sample from the dataset can be seen in Figure 1. For our task, we selected 38 popular artists with different styles and from various decades (Figure 2).

3 Authorship attribution

3.1 Description

3.2 Pre-processing

3.3 Feature extraction

3.4 Models

3.5 Feature work

4 Emotion analysis

Anca - Lexicon + cum procesezi + explicatii sentiment analisys

	artist	year	song	album	lyrics
0	ABBA	1973.0	Another Town, Another Train	"Ring Ring"	\n\nDay is dawning and I must go\nYou're aslee...
1	ABBA	1973.0	Disillusion	"Ring Ring"	\n\nChanging, moving, in a circle\nI can see y...
2	ABBA	1973.0	People Need Love	"Ring Ring"	\n\nPeople need hope, people need loving\nPeop...
3	ABBA	1973.0	I Saw It In The Mirror	"Ring Ring"	\n\nI saw it in the mirror, I saw it in my fac...
4	ABBA	1973.0	Nina, Pretty Ballerina	"Ring Ring"	\n\nEvery day in the morning on her way to the...

Figure 1: Samples from the dataset

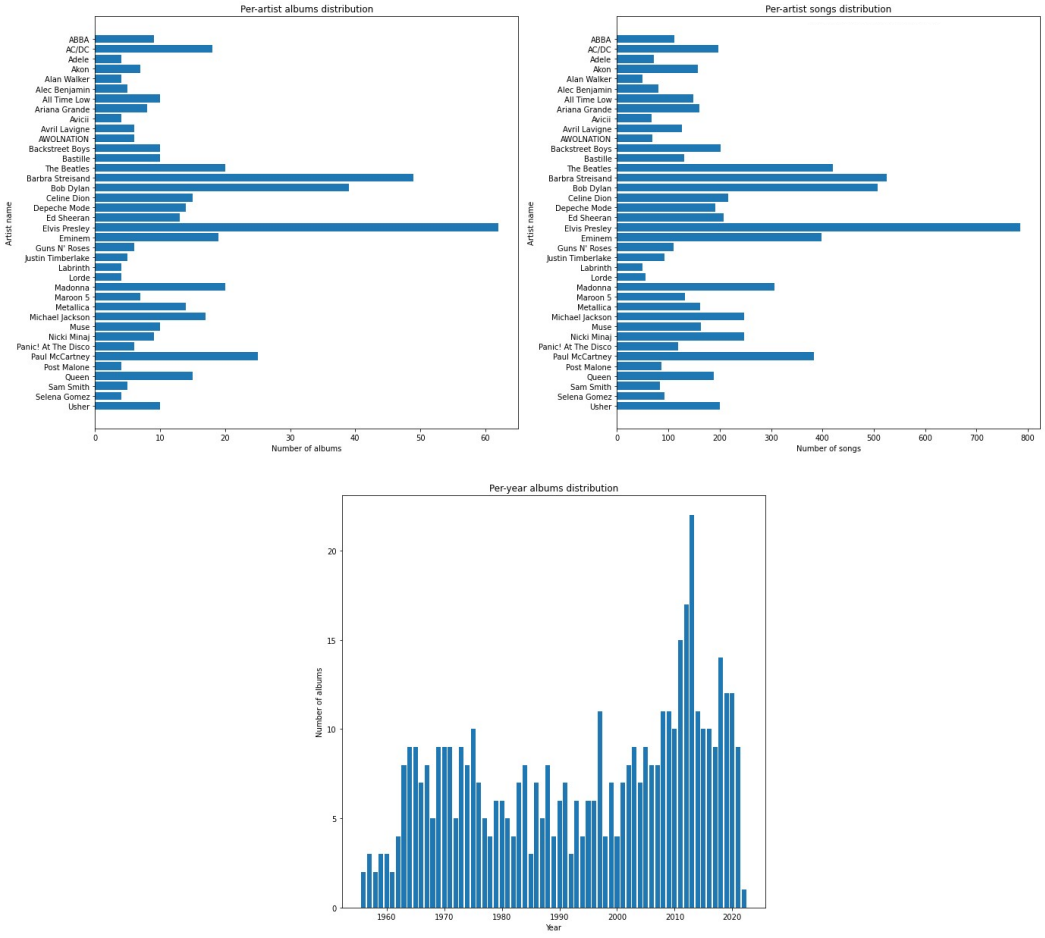

Figure 2: The number of songs per artist

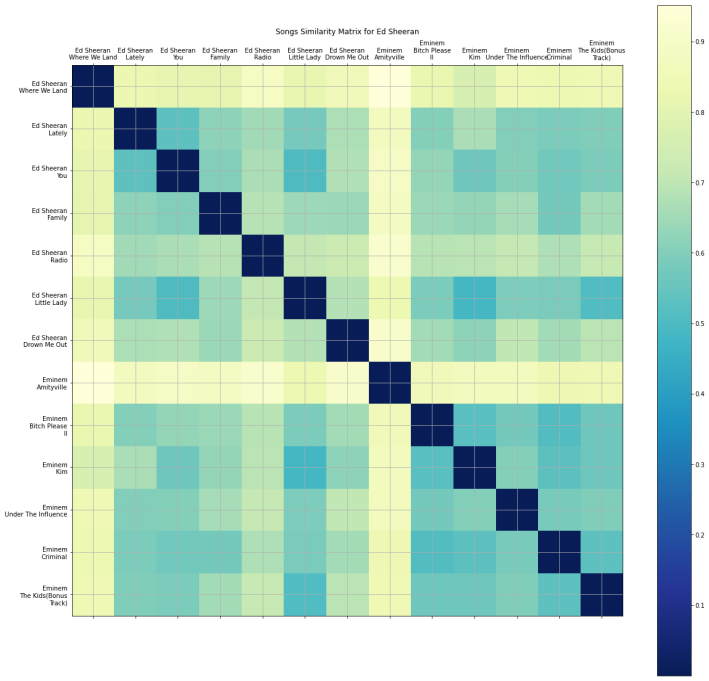

Figure 3: Song similarity between the songs of Eminem and Ed Sheeran

4.1 Defining emotions

4.2 The Emotion Lexicon

4.3 Text processing

4.4 The analysis algorithm

4.5 Results and statistics

4.6 Possible improvements

5 Song Similarity

In order to the compute song similarity, we transformed the lyrics with TfidfVectorizer and then applied the pairwise similarity on the resulting matrix. Better results were noticed with the raw lyrics (no preprocessing). In Figure 3 we can observe how similar Eminem's and Ed Sheeran's songs are.

5.1 Future work

The similarity algorithm might improve by using doc2vec (genism.models) instead of TfidfVectorizer since the meaning of the words would be taken into consideration. Another change would be to experiment with different functions such as cosine similarity (sklearn.metrics.pairwise).

References

[AZL] AZLyrics. Azlyrics - song lyrics from a to z.

[Bea] BeautifulSoup. Beautiful soup documentation.