Predicting lyrical sentiment through multivariate linear regression on audio features

Luitwin Mallmann[a], David P Serfaty[b], Evangelos Tricoupis[c] and Max Theodorus Maria van der Werf[d]

Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, Netherlands

Abstract

This paper aims to introduce the text mining technique of sentiment analysis in combination with audio feature extraction and the Spotify API to investigate the relationship of musical features and its relationship with the emotions embedded in the lyrics of a song. For this purpose, multivariate regression with audio features derived from a Spotify playlist of 1111 songs (after filtering 738 songs) are used. These are mode, valence, energy, liveness and danceability as independent variables and a sentiment score from the lyrics of these songs as dependent variable. The paper begins by illustrating the data mining concept, the main musical concepts and research that we make use of which are sentiment analysis itself, sentiment analysis in musicology and the musical concept of a music text relationship. Moreover, the audio features and their expected relationship with the lyrics sentiment score are outlined. After reviewing the methodology this paper presents the results. The paper finds that none of the 5 hypotheses could be confirmed. Furthermore, the results show the variable danceability to be significant but surprisingly negatively correlated with the sentiment score of lyrics. Also, the variable popularity which was controlled for was found to be positively significant. After discussing whether these findings agree with our expectations and the findings of other scholars, the paper makes suggestions for future research or different approaches that could hav yielded a better fitting model. Lastly, the limitations of this analysed are presented. Although this paper is strongly associated with musicology we aim to present the findings in a way that they are equally comprehensible for a person familiar with data science methods without specific musical background knowledge.

Keywords

sentiment analysis, Spotify, lyrics, linear_model

1. Introduction

The connection between music and lyrics has likely been looked at since the existence of song interpretation. A major part of musicology attempts to identify how lyrics can amplify the feelings of a song and vice versa. This has and oftentimes is still done by looking at individual parts of the lyrics and the corresponding musical features at the corresponding point in the song. The relationship between the music and the text of

l.f.r.mallmann@students.uu.nl (L. Mallmann); d.p.serfaty@students.uu.nl (D.P. Serfaty); e.tricoupis@students.uu.nl (E. Tricoupis); m.t.m.vanderwerf2@students.uu.nl (M.T.M.v.d. Werf)

a song can however also be analysed on a broader level by looking at certain elements that apply to the song as a whole such as tempo or mode of a song as well as the overall sentiments of the lyrics of a song. It can be a very tedious process when one attempts to first extract the aforementioned elements from a large corpus of songs as one would have to find lyric sentiments and secondly then determine the musical features for each song.

To tackle the first task, sentiment Analysis will be carried out to extract text from lyrics. In regards to the second task, the Spotify API provides data on over 13 audio features such as valence, danceability, key, mode or duration of the song. This will be used to construct variables for musical elements. With the combination of these two techniques, this paper aims to investigate on the relationship of musical features and its relationship with the emotions embedded in the lyrics of a song. By this, the paper adds aims to add novel findings to the existing musicology research. Various researchers have analysed song lyrics using Sentiment Analysis (Napier, 2018 etc)[2]. Scholars have also scholars connected these sentiments with musical features(Sun et al, 2018)[3]. Their methodology however uses different musical features, as well as a different sentiment analysis technique,to predict sentiment which is why this paper finds its relevance among the existing research. This paper starts by elaborating on the musical concept of the relationship between music and text and the different audio features that will be used to predict lyrical sentiments. From this, 5 hypotheses regarding the influence of audio features on the lyrical sentiment are outlined. Following an overview of sentiment analysis are the methods carried out through this research. After that, the paper presents the multivariate regression analysis with and discusses its relevance in the context of musicology. The conclude our findings as well as the limitations of our analysis.

2. Literature Review

2.1. Sentiment Analysis

Natural language processing (NLP), text analytics and computational linguistics are the applications referred to in Sentiment Analysis that are used to identify and extract information in source materials (Kaur et al, 2013)[4]. Human-Computer Interaction (HCI) practitioners and researchers, sociologists, marketeers, psychologists, economists and political scientists tackle a wide range of problems using Sentiment Analysis. (Gilbert, 2014)[5]. It is expected that Sentiment Analysis can also be applied to the scope of musicology with the main concept being the labelling of the polarity of song lyrics, in regards to positive (good), negative (bad) or neutral (surprise) emotions. In this paper sentiment will be represented by the variable compound which is calculated with the following formula which is based upon the Vader package and dictionary for sentiment analysis:

$$\frac{x}{\sqrt{x^2 + a}} \tag{1}$$

"To calculate the sentimental score of the entire text, Vader scans the text for known sentimental features, modified the intensity and polarity according to the rules, summed up the scores of features found within the text and normalized the final score to (-1, 1) with the above-mentioned formula." (Ma, 2020). Vader also gives us the percentage of positive, negative and neutral emotions, which will be used in the results section.

2.2. Sentiment Analysis in Musicology

A recent study by Napier et al. has demonstrated how sentiment analysis can be utilized to analyse the sentiments of musical lyrics over time and how they have changed. By performing automatic sentiment analysis on 6,150 on all Top 100 Billboard songs from the years 1951 until 2016, it was shown that there was a clear and statistically significant change in sentiments of popular songs expressed through lyrics and generally towards a negative tone (Napier et al., 2018)[2].

2.3. Sentiment Analysis and Music

Another study by Sun et al. looked at the connection of the lyrics with other musical features which also the aim of this paper. They analysed how lyrical emotions are expressed in music, by exploring the correlation between affect-carrying lyrics and musical features, some of those being beat strength, duration, pitch height, consonance, and mode. The study revealed, "that metrical strength and note lengths are highly correlated with affects, while correlations of pitch height, consonance, and mode are in general less significant, at times contradicting previous research" (Sun et al, 2017) [3].

Whereas the aforementioned study analyses the music-text connection on the individual note level, the work presented in this text is excusively based upon audio data extracted from Spotify for the entirety of a single song. The next section will outline what is meant by the relationship between music and text and how the musical features for this paper are defined, as well as how they are expected to be correlated with the lyrical sentiments.

3. Key Concepts and Hypothesis

3.1. Music-Text Relationship

The research aims to support or reject a relationship between the sentiments of the lyrics and the sentiments of music. The idea behind the concept of music-text relationship is that the text supports the music and vice versa. An example can be drawn out, where an artist sings the words "high" and at the same time the melody of the song goes to a very high note. Unlike this manually annotated and isolated example, text mining techniques can be employed to analyse the music-text relationship on a broader level given the audio features extracted from Spotify and the sentiment analysis of the lyrics which are always a summary of values for the song as whole. It is expected that certain audio features will be especially correlated with the sentiments of the lyrics. The following subsections will lay down a foundation of music theory in regards to the audio features to be extracted.

How these are expected to be related to the sentiments of the lyrics of a song, and their importance and relevance to the research will follow.

3.2. Mode

As noted by the Spotify API, mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived [6]. Major is represented by 1 and minor by 0.

In other words, songs in minor are usually associated with a sad and often melancholic mood whereas songs in major are mostly associated with happiness. Therefore, the mode (minor or major) is something that is expected to have a strong relationship with the songs' lyrics. Consequently, songs in minor are expected to have more negative sentiments score and songs in major are expected to have more positive sentiments score. Sun et al. found no support that minor mode pieces made more use of negative and sad words than major mode pieces, it will be interesting to see if the following hypothesis is also insignificant with our dictionary and sentiment and audio extraction technique and positiveness conveyed by a track which will be tested with the following hypothesis:

H1: There is a significant relationship between the mode of a song and its lyrics sentiment score.

3.3. Valence

The next audio feature is valence, represented by a float between 0 and 1. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry) [6]. This is a variable that is expected to encompass the overall musical emotions a song and it is expected that it will be the most significant of all variables. As a consequence, this variable should have the strongest relationship with the lyric sentiments of a song. The following hypothesis shall be put to the test, to assess this relationship.

H2: There is a significant positive relationship between the valence of a song and its lyrics sentiment score.

3.4. Danceability

This next feature assigns a value between 0 and 1, describing how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 indicates a song that is least danceable and a value of 1.0 indicates a song that is most danceable [6]. Dancing is often related to "carefreeness" and "leaving worries behind". As with all types of exercise, dancing induces the activation of chemicals, particularly dopamine, oxytocin, serotonin, and endorphins which play a major part in triggering positive emotions, such as pleasure, happiness and love [7]. Intuitively, this should also become visible in a song's lyrics. Sun et al. found that "notes carrying texts with negative sentiment are consistently on weaker beats than words with other effects or words in the NRC vocabulary with neutral affect". Although beat strength is not the same as danceability, some correlation is expected since it is part of how danceability is constructed.The variable "Danceability" as provided by Spotify has not been used in other studies which helps pose interesting questions. With this, the following hypothesis is constructed.

H3: There is a significant positive relationship between the danceability of a song and its lyrics sentiment score.

3.5. Energy

The feature "energy" is assigned a value between 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Spotify notes that energetic tracks feel fast, loud and noisy [6]. While this a rather broad concept, it is expected that this feature can capture the enthusiasm of a song well. Albeit probably than the other feature, this variable could be still very well be correlated with the sentiments of a song's lyrics which leads to the following hypothesis.

H4: There is a significant positive relationship between the Energy of a song and its lyrics sentiment score

3.6. Liveness

"Liveness detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides a strong likelihood that the track is live." This relationship might seem a bit less obvious. Nonetheless, we expect that the reaction of an audience can indicate that a song transmits a rather happy message. Intuitively, we derived that an audience is less likely to respond loudly to a song especially if they are aware of the sad message the lyrics reflect. Although most likely the least significant of all, but interesting due to its novelty in studies, we derive the following hypothesis.

H5: There is a significant positive relationship between the Liveness of a song and its lyrics sentiment score.

4. Methods

This section dives into the analysis of how this paper goes about analysing music text relationship using a Spotify playlist compiled for this study. The entire analysis is done via Python from the data extraction to the analysis. The playlist the analysis is based upon contains 1111 songs. Using the Spotify API, the playlist is transformed into a dataset that contains the following variables:

- track_name
- popularity
- danceability
- length
- energy
- loudness
- speechiness
- acousticness
- instrumentalness
- liveness
- valence
- tempo
- duration
- time
- signature
- label

The Spotify API [6] outlines the meaning for each of these variables. This research shall focus on the variables that were elaborated upon in section 3, from which the hypotheses were constructed. Next to the audio variables clarified in hypothesis, 6 control variables will be put in place such as popularity and duration to control for non expected correlations. Moreover, the lyrics for each song were extracted. The lyrics of the curated Spotify playlist are extracted using the Genius API, a platform that provides lyrics for a large number of songs[8].Then the lyrics with the data frame of the audio feature data were combined. The independent variable *compound* is then constructed based upon the extracted lyrics with the aforehand mentioned formula. The sentiment analyzer function of the NLTK Python library is used to calculate the positive, neutral and negative sentiment scores. The sentiment analysis conducted was based on the VADER dictionary, a simple rule-based model for general sentiment analysis. This dictionary has been proved superior effectiveness compared to other lexicons (Gilbert, 2014)[5]. Using pandas, the songs for which no lyrics could be found were dropped. Although the analysis combined 9 top 100 charts from the US, 66 non-English songs were present in these charts which we filtered out using a language detection model. We then end up with our final data frame which is shown below of 738 songs.

	track_name	artist_x	lyrics	neg	neu	pos	compound	album	artist_y	release_date	valence	length	popularity
0	Hold On	Kamaal Williams	[...]	0.13	0.838	0.031	-0.8702	Wu Hen	Kamaal Williams	24/07/2020	0.146	200596	41
1	What Kinda Music	Tom Misch	[...]	0	0.97	0.03	0.3818	What Kinda Music	Tom Misch	24/04/2020	0.151	230470	57
2	Do It	Chloe x Halle	[...]	0.032	0.811	0.157	0.9938	Ungodly Hour	Chloe x Halle	03/06/2020	0.566	176786	73
3	JEWELZ	Anderson .Paak	[...]	0.056	0.765	0.179	0.9466	JEWELZ	Anderson .Paak	06/10/2020	0.766	174213	69
4	Shades of You	Moses Boyd	[...]	0	0.805	0.195	0.9483	Dark Matter	Moses Boyd	14/02/2020	0.727	260122	38
5	Pretty Please	Dua Lipa	[...]	0.09	0.64	0.27	0.981	Future Nostalgia	Dua Lipa	27/03/2020	0.656	194607	74
6	Plastic Plants	Mahalia	[...]	0.054	0.761	0.185	0.9766	Isolation Tapes	Mahalia	01/05/2020	0.752	199533	56
7	Cabin Fever	Jaden	[...]	0	0.902	0.098	0.9146	Cabin Fever	Jaden	23/07/2020	0.917	195535	65
8	Rose Rouge	Jorja Smith	[...]	0.054	0.76	0.186	0.8902	Rose Rouge	Jorja Smith	09/06/2020	0.299	357960	62
...	...	...	...	...	...	...	...	...	...	...	...	...	...
731	St. Eriksplan	Low Roar	[...]	0.055	0.789	0.156	0.8973	Once In a Long, Long...	Low Roar	14/04/2017	0.194	221187	35

Table 1: first half of dataframe

	danceability	acousticness	energy	instrumentalness	liveness	loudness	mode	tempo	time_signature	valence	speechiness	analysis_url	language
0	0.304	0.963	0.243	0.154	0.103	-15.443	1	102.898	3	0.146	0.0408	[...]	en
1	0.523	0.0532	0.56	0.83	0.0981	-9.423	0	96.167	4	0.151	0.145	[...]	en
2	0.725	0.0932	0.668	2.33E-06	0.112	-7.286	0	82.966	4	0.566	0.055	[...]	en
3	0.884	0.371	0.653	0.00135	0.812	-7.653	0	108.548	4	0.766	0.0586	[...]	en
4	0.656	0.0812	0.956	0.00759	0.249	-6.338	0	126.992	4	0.727	0.104	[...]	en
5	0.906	0.0311	0.474	5.52E-06	0.286	-6.124	1	106.976	4	0.656	0.194	[...]	en
6	0.823	0.636	0.441	7.68E-05	0.114	-7.645	1	135.947	4	0.752	0.156	[...]	en
7	0.729	0.134	0.874	0.00011	0.198	-5.585	1	123.001	4	0.917	0.0324	[...]	en
8	0.663	0.187	0.703	0.326	0.319	-12.125	0	117.953	4	0.299	0.0562	[...]	en
...	...	...	...	...	...	...	...	...	...	...	...	...	...
731	0.486	0.731	0.236	0.601	0.12	-12.606	0	98.062	4	0.194	0.0278	[...]	en

Table 2: second half of dataframe

On the data, a multivariate regression model with the audio features as predictor variables and “sentiment compound” as the outcome variable is fitted. We then evaluate this model using R squared and adjusted R squared, and the significance of each dependent variable. To exclude to that the different variables cancel each other out, the analysis builds up the model incrementally with variables which we expect to have the most significance music-text relationship which is "valence", then we add danceability, thirdly liveness and energy and lastly our control variables until we end with our final model.Doing this the significance levels of each variable at each stage of adding new variables have however not changed a lot which tells us that the different variables do not cancel each other out. Thus, the following section can present the final model.

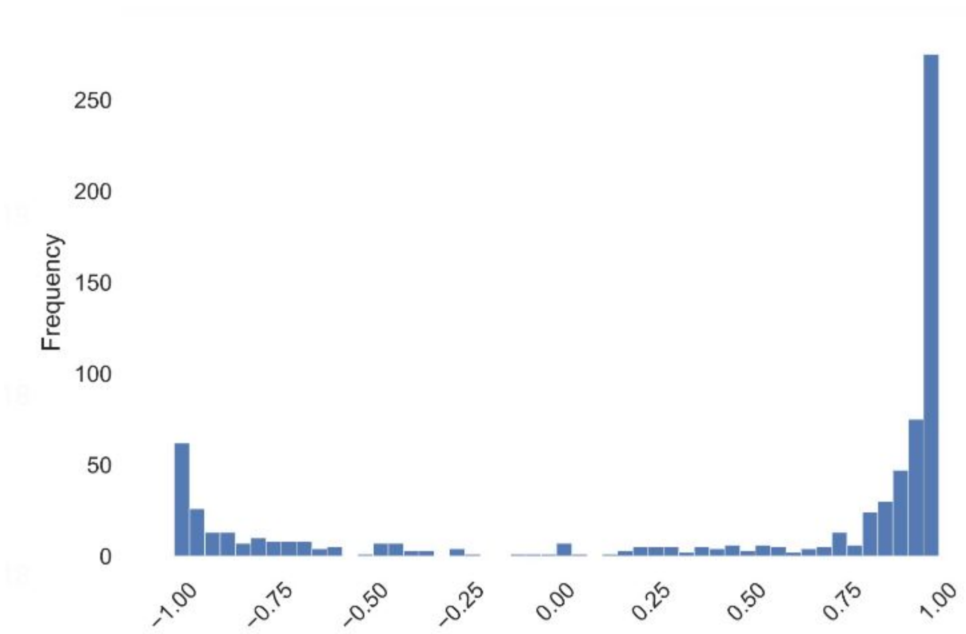
5. Results

The following table shows the results of the multivariate linear regression.

	names	coef	se	T	pval	r2	adj_r2	CI[2.5%]	CI[97.5%]
0	Intercept	0.028514	0.773059	0.036885	0.970587	0.022323	0.007387	-1.489206	1.546233
1	mode	-0.007093	0.057800	-0.122720	0.902363	0.022323	0.007387	-0.120570	0.106384
2	valence	-0.006350	0.160327	-0.039605	0.968419	0.022323	0.007387	-0.321115	0.308415
3	danceability	-0.680989	0.253540	-2.685927	0.007400	0.022323	0.007387	-1.178754	-0.183224
4	energy	-0.018519	0.270001	-0.068589	0.945336	0.022323	0.007387	-0.548603	0.511565
5	liveness	0.065486	0.214656	0.305075	0.760398	0.022323	0.007387	-0.355940	0.486912
6	speechiness	0.238396	0.358516	0.664953	0.506293	0.022323	0.007387	-0.465465	0.942257
7	instrumentalness	0.243252	0.312789	0.777686	0.437009	0.022323	0.007387	-0.370836	0.857339
8	popularity	0.002005	0.000936	2.143292	0.032424	0.022323	0.007387	0.000168	0.003842
9	loudness	0.015275	0.021422	0.713043	0.476050	0.022323	0.007387	-0.026782	0.057331
10	tempo	0.000051	0.001170	0.043776	0.965095	0.022323	0.007387	-0.002246	0.002349
11	time_signature	0.204553	0.178965	1.142973	0.253430	0.022323	0.007387	-0.146804	0.555909

Table 3: Multi-variate regression results

The variable danceability is one of two significant variables. It is strongly significant at the 1 per cent level. It is also negative with a coefficient of -0.68 suggesting that the more "danceable" a song is, the less happy it is. From the control variables, only *popularity* was significant however only slightly positive with a coefficient of 0.002.
After checking the assumptions for multiple regression, it became apparent that our dependent variable violated the assumption of normal distribution. The majority of songs have very happy lyrics or sad songs and there are very few neutral songs.
The distribution of this variable is shown in the following graph:



Distribution of the outcome variable "compound"

The results should be interpreted accordingly. To compensate for this, a new variable was created based on the negative and positive sentiments of a song where we simply deduct the negative score from the positive sentiment score. This new variable was normally distributed and with it in place, regression results were however still similar only with Danceability being less significant as shown below:

	names	coef	se	T	pval	r2	adj_r2	CI[2.5%]	CI[97.5%]
0	Intercept	0.031529	0.115596	0.272754	0.785121	0.019938	0.004965	-0.195417	0.258476
1	mode	-0.005508	0.008643	-0.637275	0.524148	0.019938	0.004965	-0.022476	0.011460
2	valence	0.009194	0.023974	0.383505	0.701459	0.019938	0.004965	-0.037873	0.056261
3	danceability	-0.072833	0.037912	-1.921109	0.055113	0.019938	0.004965	-0.147265	0.001598
4	energy	0.006747	0.040374	0.167118	0.867324	0.019938	0.004965	-0.072517	0.086011
5	liveness	0.004504	0.032098	0.140329	0.888440	0.019938	0.004965	-0.058512	0.067521
6	speechiness	-0.000226	0.053609	-0.004223	0.996632	0.019938	0.004965	-0.105476	0.105023
7	instrumentalness	0.082685	0.046772	1.767836	0.077512	0.019938	0.004965	-0.009140	0.174510
8	popularity	0.000299	0.000140	2.138765	0.032791	0.019938	0.004965	0.000025	0.000574
9	loudness	0.003589	0.003203	1.120286	0.262965	0.019938	0.004965	-0.002700	0.009877
10	tempo	0.000032	0.000175	0.180307	0.856962	0.019938	0.004965	-0.000312	0.000375
11	time_signature	0.017774	0.026761	0.664169	0.506795	0.019938	0.004965	-0.034765	0.070312

Table 4: Multi-variate regression results with the updated variable

Looking at the new results, only *instrumentalness* is slightly significant.

6. Discussion

No hypotheses can be confirmed. Based on the results it cannot be confirmed that a song's *mode*, *valence*, *energy* or *dancability* are positively correlated with a positive sentiment score on the lyrics of the song. Our findings do, nonetheless, not entirely disagree with other scholars that investigated the relationship of music of emotion in lyrics. Although negatively correlated, the paper found that "Danceability" is significant in predicting a song's lyric sentiment. Sun et al. found that metrical strength is highly correlated with a lyric's sentiment. Since beat strength is also part of how the variable danceability is computed (spotify.com), an overlap in this aspect is likely. Moreover, similar to Sun et al., we also detected very little significance in the mode (major or minor) of a song's lyrics. Apart from that it was rather surprising to find that the variable valence which aims to capture the musical positiveness conveyed by a track was very insignificant in predicting the dependent variable. Analysing this discrepancy is subject for a different research question. Also, the variable energy was insignificant in predicting the variable compound. The same can be said for the variable liveness which depicted the presence of an audience in a recording. Since our expectations for the relationship of energy and liveness with the compound was purely derived from intuition and logical reasoning, the results are not necessarily surprising. Since to our knowledge, the relationship of sentiments with these two has not evaluated by other scholar it would have been interesting to find significant findings. Overall, it can be said that lyric emotions remain hard to predict based solely on audio features and the controlling variable popularity.To improve the model, one could also include genre of a song which can also be derived using the Spotify API. In that case one would also have to have a more diversified musical spectrum than predominantly pop songs which we used due to expected higher availabilty of lyrics and playlist. To find more significance, another strategy could also be to not only differentiate between positive and negative emotions, but to break these down into more precise emotions such as anger, melancholia, love which however be more of a mood classification analysis. Also, with more advances in sentiment analysis that can detect emotion in sentences as a whole or a multiple words, more sophisticated song data also on the local level of a song, one could also expect the model to improve.

7. Practical Implications

The model tested shows very low r^2 and adjusted r^2, that is, the variation in the output variables can barely be explained. As such, the findings suggest that the model could for example not be used by a recommender system at Spotify. More precisely, if the model were to perform satisfactorily, it could be implemented in Spotify or other streaming plattform with similar audio feature data, to make recommendations based upon these audio features to for instance recommend sad lyrical songs if there are no lyrics available by the streaming platform itself.

8. Conclusion

The model created during this research was not successful at predicting the emotions of a song's lyrics when using audio features from the Spotify API. Nevertheless, this paper could demonstrate successfully how sentiment analysis using the VADER dictionary can be applied in musicology as well as how Spotify audio features can be used for this kind of analysis. This was the main aim of this paper given that it is part of the data mining course and not a musicology course. Next to that, the regression result showed that danceability negatively influences the sentiment score of a song's lyrics and the control variable popularity significantly, but only slightly positively impacts the emotions of a song's lyrics. All the other variables namely valence, energy and liveness were not significant at explaining the relationship. Although the model that was generated might not be very practical, it could be improved by including other musical variables as well through advancements in the sentiments analysis. With the dataset that we generated other research questions can be investigated. For example, how the sentiments of the lyrics correlate with the popularity of a song.

9. Limitations

Firstly, it should be mentioned that we evaluated a music-text relationship based upon averaged musical features of an entire piece. These techniques are not able to elaborate on one distinct text-music relationship for example singing the word "high" with a high melody or pitch. Secondly, it is important to point out that although most of the songs nowadays convey an overall rather happy or sad feeling, music lives off the interplay of happiness and sadness. It is what created tension. Our audio features could not look at these interplay, but only the overall emotions conveyed. Lastly, we exclusively looked at English language songs since the sentiment analyser can only process English lyrics. Yet, it could also be interesting to see if our findings can be confirmed with other languages and perhaps non-western music and generally other music genres.

References

[1] Ma, Y. (2020, February 05). NLP: How does NLTK.Vader Calculate Sentiment? Retrieved January 25, 2021, from https://medium.com/ro-data-team-blog/nlp-how-does-nltk-vader-calculate-sentiment-6c32d0f5046b

[2] Napier, K., & Shamir, L. (2018). Quantitative sentiment analysis of lyrics in popular music. Journal of Popular Music Studies, 30(4), 161-176..

[3] Sun, S. H., & Cuthbert, M. S. (2018). Emotion painting: lyric, affect, and musical relationships in a large lead-sheet corpus. Empirical Musicology Review, 12(3-4), 327-348.

[4] Kaur, A., & Gupta, V. (2013). A survey on sentiment analysis and opinion mining techniques. Journal of Emerging Technologies in Web Intelligence, 5(4), 367-371.

[5] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 8, No. 1).
[6] Mode, Valence, Dancability, Energy - AudioFeaturesObject
[7] Greater Good Science Center - Dance for your DOSE of Happiness and Health.
[8] The Genius API