Computational Social Sciences
Master in Data Science
A.Y. 2020/2021

Mining patterns from songs banned after 9/11

Hrovatin Lucia
223027
lucia.hrovatin@studenti.unitn.it

Abstract

The 9/11 attacks made a substantial impact on public opinion, leading to political and societal consequences. Among them, censorship of *inappropriate* entertainment products has been introduced by suggesting to radio stations not to play songs whose lyrics were deemed to trigger semantic connections with the events. This phenomenon, called associative priming, usually interests language production, but it can be applied to music too. However, music relies on two levels of communication: lyrics and audio. In the following paper, both will be investigated using a multi-level approach. Whereas the former component is evaluated by means of Latent Dirichlet Allocation (*LDA*), the latter is retrieved from Spotify and analysed employing Random Forest. Results show how the song classification as *inappropriate* was rather driven by lyrics content than audio features. Furthermore, Random Forest performs well in discriminating censored songs but relies on features having a neutral connotation such as *danceability.*

1 Introduction

Music can be interpreted as spreader of messages [1] and trigger of emotions [2]. Moreover, if music is a song, it employs two different channels: tune and lyrics. This duality allows songs to convey semantically meaningful concepts [3] and trigger priming recognition effects [4], similarly to the linguistic associative priming [5]. Whereas the capability of arousing emotions and retrying autobiographic memories [6] empowers music to be effective in several contexts (e.g., psychotherapy [7], advertising [8]), it may show a drawback when the established connection leads to negatively connoted events. This may be the case of the 9/11 attacks and the list of *inappropriate* songs that have been temporarily banned from American radio stations [9]. Even though it was a suggestion and not an outright ban on the songs [10], the list's publication brought their broadcast to a halt. Further considerations regarding the legitimacy of the list may offer interesting insights but could also give rise to a debate about the controversial issue of censorship [11, 12]. However, it goes beyond the scope of this project and will not further be discussed. Indeed, this project focuses on the shared features of the censored songs and aims at detecting the criterion used to outline the song (in)appropriateness.

A last premise regards the general framework and the psychological impact of the attacks. Indeed, the attacks had psychological repercussions ranging from clinical mental diseases on directly exposed individuals (e.g., *Posttraumatic stress disorder, PTSD* [13, 14]) to widespread coping mechanisms to mitigate conscious and unconscious death-related thoughts [15]. Accordingly, the songs' censorship may be deemed a large-scale coping mechanism to avoid Involuntary Semantic Memory retrieval (*ISM*) [16].

In the following sections the multimodal songs' features will be processed using a multi-level clustering and classification model.

2 Previous studies

The existing research stresses the existence of a shared mechanism between language and music for domain-general rate processing [17], while two distinct brain regions are responsible for processing domain-specific syntactic rules [18, 19]. Music, in terms of lyrics, does not own formal syntactical structures but gathers a fragmented collection of words, phrases, and idioms [20]. The items within the collection have two relevant characteristics: not uniqueness (i.e., repetitions of words across songs are allowed) and (basic) semantic preservation. The presence of semantic allows an interpretation of the items (i.e., words) as nodes in a (semantic) network, within each node is linked with other words or concepts (e.g., thoughts, memories). Moreover, the network should activate via *spreading activation* [21], meaning that few words (i.e., *primers*) go beyond a certain threshold and trigger other concepts. The outcome is an elicitation of ISM and Involuntary autobiographical memory (IAM) [22]. As stated above, music conveys information that could be associated with both emotional and semantic information, either directly (i.e., lyrics) or indirectly (i.e., audio features) [23, 24]. Thus, also audio features can be defined primers because they can trigger emotions and, indirectly, evoke ISM or IAM. Accepting the underlying assumption that music acts as two levels primer, two interrelated models must be considered: categorical and dimensional [20]. While the categorical model uses human language for labelling categories and focuses on semantic information, the dimensional model describes audio component as a multidimensional vector space having a dimension per audio feature. Similar approaches have been already applied in other classification tasks, such as music emotion classification [20] and genre classification [25, 26]. More specifically, previous studies

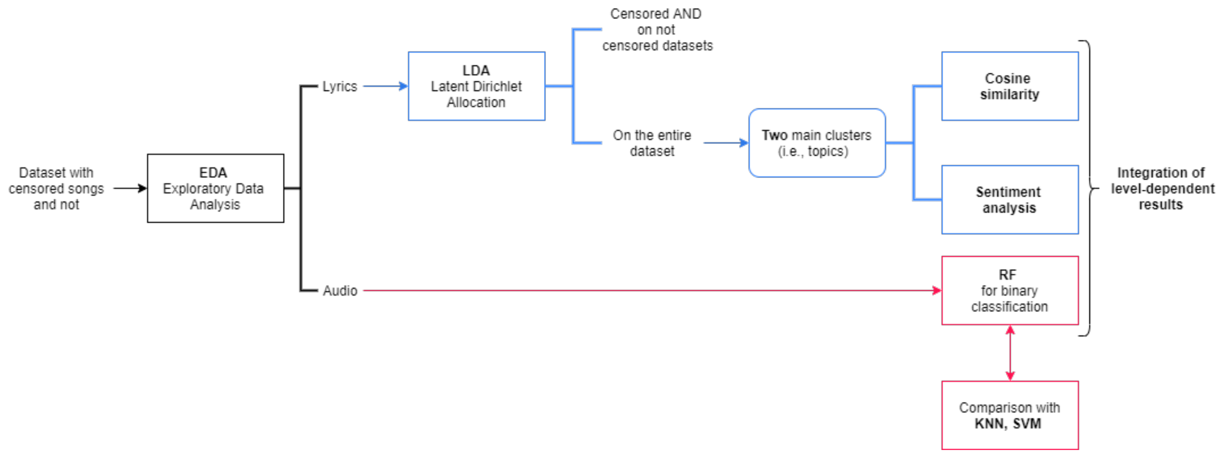

Figure 1: Multi-level model approaching the features embedded in a song: lyrics and audio.

processed lyrics and relative metadata using traditional classifiers such as Support Vector Machines, Naïve Bayes, Graph-based methods, and classification trees [27]. These methods did not show an outstanding performance and have been mainly replaced by the unsupervised algorithm for topic modelling: *Latent Dirichlet Allocation* (LDA) [28]. Music classification and emotion recognition cluster words within lyrics based on topics, but a quantitative measure of (dis)similarity among topics is achieved by running additional algorithms, such as cosine similarity [27].

According to the multi-level model previously stated, vectors of audio features represent the dimensional model. In the existing literature, those features are extracted either directly, formatting the tune according to specific principles [26], or indirectly, parsing the data already stored by music streaming services such as Spotify [29, 30]. Thus, these services offer a wide on-demand catalog of digital content and also collect and store customized music (meta)data, allowing the music information retrieval (MIR) [31] for real world applications. Examples are user recommendations [32], genre classification [25], hit prediction [29], and emotion recognition [33]. Audio features have been employed in several classification tasks and the statistical learning model of Random Forest seems to outperform the others [29]. The remainder of this paper aims to detect patterns able to justify censoring those songs after the 9/11. Specifically, two research questions are posed: (1) were the choices based on lyrics semantic (i.e., direct primer), or did they also consider audio features (i.e., indirect primer)? (2) If any pattern is found, should it be generalizable to other catastrophes or events?

3 Methodology

3.1 Dataset

Any existing dataset was suited for the following analyses; therefore, data has been gathered and integrated to contain all the multimodal music features [1]. Firstly, each song on the list distributed by Clear Channel[2] has been found on the music streaming service Spotify[3] and saved in a publicly available playlist[4]. Via the Spotify Web API[5], unique identifiers (e.g., track id, artist id), general information (e.g., track title, artist name), audio components, and other features processed by Spotify (e.g., popularity) have been downloaded and stored on a local file. Neither data nor metadata coming from Spotify met the second demand, meaning the lyrics of the songs. Therefore, their acquisition involved the Genius.com API, wrapping data stored in the Genius.com[6] collection, and a Python parser for storing the texts and the corresponding Spotify track unique identifier. For balancing data, 100 tracks randomly taken from the Billboard hits ranking of 2000[7] and among the most popular songs of the 90s[8] have been gathered applying the same dual-step procedure as above[9]. The final dataset includes 283 entries specified by the following 12 attributes:

- *track_ id*: the song's unique Spotify track identifier.
- *artist_ id*: the artist's unique Spotify identifier.
- *danceability*: value ranging from 0.0 to 1.0 and describing the suitability of the track for dancing. It combines musical elements such as tempo, rhythm stability, beat strength, and overall regularity[10].
- *energy*: a perceptual measure of intensity and activity. Its values range from 0.0 to 1.0.

[1] The explanation of the data collection procedure is also attached to the code folder as READ.ME file

[2] https://en.wikipedia.org/wiki/Clear_Channel_memorandum

[3] https://www.spotify.com/

[4] https://open.spotify.com/playlist/61AzVorsMoZq3vSYQ01L1y?si=0e116d6c3d8e48fb

[5] https://www.rcharlie.com/spotifyr/

[6] https://pypi.org/project/lyricsgenius/

[7] https://en.wikipedia.org/wiki/Billboard_Year-End_Hot_100_singles_of_2000

[8] https://www.rollingstone.com/music/music-lists/50-best-songs-of-the-nineties-252530/

[9] https://open.spotify.com/playlist/3Ir71PTD9niR2MFLGwi8Kg?si=70ab23da70704b6e

[10] https://developer.spotify.com/documentation/web-api/reference/

- *key*: is the major or minor scale around which a track revolves.
- *loudness*: the overall loudness of a track in decibels (dB).
- *speechiness*: a value ranging from 0.0 to 1.0 describing the number of spoken words.
- *valence*: describes the positiveness of a track and ranges from 0.0 to 1.0. Values closer to 1.0 suggest a more upbeat track[11].
- *instrumentalness*: a value from 0.0 to 1.0 measuring whether the track is instrumental or contains vocals.
- *acousticness*: a value from 0.0 to 1.0 predicting whether the track is acoustic.
- *popularity*: a value between 0.0 and 1.0, where values closer to 1.0 identify high popularity.
- *lyrics*: single string containing the lyrics of each track.

3.2 Analytical strategy

The presented strategy relies on a multi-level approach that considers lyrics and audio at two different levels and analyses them separately, as shown in the Figure 1.

3.2.1 Lyrics model

When evaluating lyrics and the conveyed messages, a necessary premise regards the subjectivity and multiplicity of moods [34] and feelings invoked by a song. Whereas these two characteristics refer to the listener's emotional state, each song also has a hidden sentimental structure based on its *topics*. When the objective corresponds to recognize and categorize songs based on topics, *Latent Dirichlet Allocation* (LDA)[28] may be entailed. LDA is a statistical approach to document modelling (i.e., lyrics) aiming at detecting latent semantic topics in a large *corpus*, meaning the collection of M documents. The basic idea behind LDA is that documents are random mixtures over latent topics, whose specificity depends on words' (i.e., basic unit of discrete data) distribution[28]. Moreover, LDA strongly asserts that each word has some semantic information, and in documents with similar topics, a similar subset of words tends to co-occur. Thus, each inferred topic is a collection of words sorted using the frequency of co-occurrences over documents. In order to quantify the (dis)similarity between inferred topics, each topic is identified as an ordered vector of n dimensions, corresponding to the top-n words [35]. The similarity between topics corresponds to the correlation between the vectors. The correlation is quantified using the cosine of the angle between vectors: the so-called cosine similarity measure. It is one of the most popular similarity measures applied to text mining tasks, such as information retrieval [36] and clustering [37]. Specifically, given two topics $\overrightarrow{T_1}$ and $\overrightarrow{T_2}$ n-dimensional vectors, their cosine similarity is:

$$SIM(\overrightarrow{T_1}, \overrightarrow{T_2}) = \frac{\overrightarrow{T_1} \cdot \overrightarrow{T_2}}{\|\overrightarrow{T_1}\| \|\overrightarrow{T_2}\|} = \frac{\sum_{i=1}^{n} w_{1i} w_{2i}}{\sqrt{\sum_{i=1}^{n} w_{1i}^2} \sqrt{\sum_{i=1}^{n} w_{2i}^2}} \tag{1}$$

where w_{1i} and w_{2i} are i^{th} components of $\overrightarrow{T_1}$ and $\overrightarrow{T_2}$ respectively.
Each component represents a word having a non-negative weight. Consequently, the cosine similarity returns a value that is non-negative and bounded between 0 and 1. A cosine value of 0 identifies two orthogonal vectors and, reversely, the closer the cosine value is to 1, the smaller the angle between vectors.

3.2.2 Audio model

The second model involves audio and Spotify features, aiming at labelling censored and not censored songs. The classification task will employ *Random Forest* (RF) [38]. RF is a statistical learning model for classification that improves the predictive accuracy of decision trees by aggregating them. In classification tasks employing Spotify data, it also outperforms other methods such as *Support Vector Machines* (SVM) and *k-Nearest Neighbours* (k-NN) [29]. Specifically, RF is an ensemble of classification trees, where each tree is built using a bootstrap sample of data and considering only m randomly chosen predictors, with $m < p$ (p = total number of predictors). Applying two levels of randomness RF decorrelates the classification trees, averages them and results a less variable and hence more reliable method [39]. The algorithm implemented in RF is similar to bagging and differs only in the choice of the hyperparameter m, the predictor subset size. Thus, $m = p$ corresponds to the bagging procedure. In the RF here used, the parameter m has been programmatically tuned using the reduction of the *Out-of-Bag error* (OOB) and of the misclassification error rate on the test set (ER), as Figure 2 shows. The optimal value corresponds to the minimum score in the test set, but it changes depending on the random seed set. In order to allow reproducible results, RF has also been run using $m = \lfloor \sqrt{p} \rfloor$, a common choice in classification setting [39].

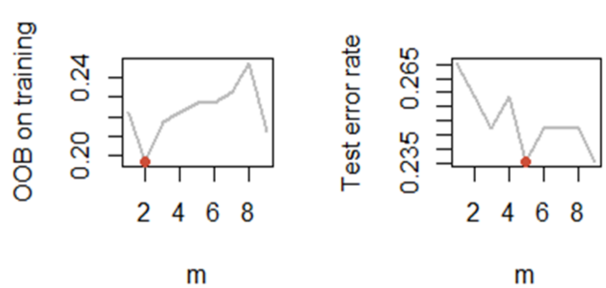

Figure 2: Example of OOB and ER performance variation depending on m, size of predictors' sample

[11] https://developer.spotify.com/documentation/web-api/reference/

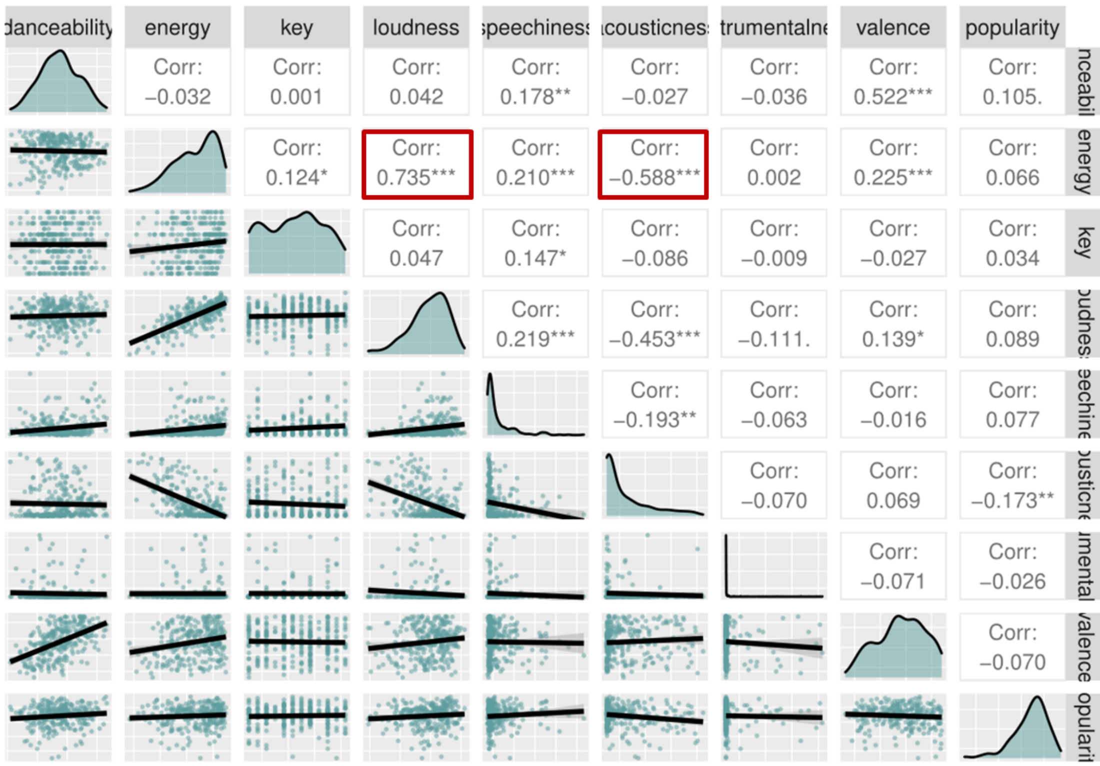

Figure 3: Pairwise scatter plots, density function for each predictor and Pearson correlation scores. The pair whose Pearson correlation coefficient is almost $\rho = \pm 0.60$ or greater are highlighted in red.

RF was chosen for some reasons. Firstly, it has a good predictive performance even when predictors contain noise or are highly correlated. Secondly, the output is invariant to monotone transformations of the predictors. Lastly, the classification trees are independent and built with two levels of randomness, avoiding any risk of overfitting [39].

$$ER_i = \sum_{j \in C_i} \frac{\mathrm{I}(y_i \neq y_j^*)}{|C_i|} \tag{2}$$

Given these characteristics and the previous literature [40], it is important to understand the performance of RF compared to alternative classifiers. The chosen classifiers are SVM and k-NN. Whereas the former solves the minimization problem and finds an optimal hyperplane to separate the songs, the second is a memory-based learner algorithm that reuses the information previously acquired and computes the Euclidean distance between the new instance and its neighbours [39]. The tuning parameter k, meaning the number of neighbours to consider, has been programmatically computed, selecting the optimal test error, hence minimizing the error rate of reported in the formula 2.

3.3 Example of analysis

This paper explores how music features cluster censored songs according to two subjects: (1) embedded topics and (2) audio features. The first step after collecting data was an Exploratory Data Analysis (EDA) to explore the main characteristics of feature distributions and possible correlations between them. The results of the Pearson correlation coefficient, the pairwise scatter plots and the distribution of each feature are summarized in Fig. 3. *Energy* and *loudness* features have a high positive correlation reaching $\rho = 0.73$, while the pair *energy* and *acousticness* shows a moderate negative correlation around $\rho = -0.59$. After this first data exploration, the multi-level model is employed and therefore the analysis follows two distinct processes.

One process involves a topic modelling analysis employing the LDA approach. The analysis is carried out on the two distinct datasets, censored songs and not, and on all collected lyrics. While the first process aims at discovering five embedded topics, the second clusters the dataset into two groups. It has been hypothesised that that the key topics' semantic of the two groups should differ. The ten most relevant terms are further evaluated in terms of cosine similarity and different sentiment analysis with three lexicons: Syuzhet, NRC and BING[12] (Fig. 4 and Table 1). Interestingly, the dissimilarity between clusters is milder than expected, reporting an overall cosine similarity of 0.82, but the sentiment expressed by the first terms, *burn* and *love*, seems to disagree with the high similarity previously declared. It is worth noticing that the difference between groups is (re)confirmed milder when the top ten terms are considered.

[12] https://www.tidytextmining.com/sentiment.html

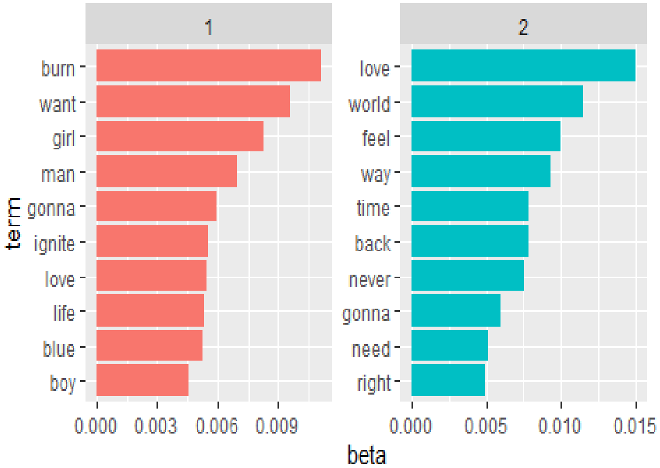

Figure 4: Frequency of top 10 words per topic

Topic	Word	Syuzhet	NRC	BING
2/1	*love*	0.75	joy/positive	positive
2	*right*	0.80	/	positive
2	*time*	/	anticipation	/
1	*burn*	−0.40	/	negative
1	*blue*	/	sadness	/
1	*boy*	0.25	disgust/negative	/

Table 1: Summary table of sentiment analysis on top 10 terms based on *Syuzhet*, *NRC* and *BING* methods

The second level turns to the experimental evidence on audio features. In this framework, the classification task is performed using a validation set approach. This approach splits the dataset into two samples, 70% as training set and the remaining as test or validation set. As previously stated, the chosen classifier is RF, whose tuning parameter m is either tuned (Fig.2) or set to $\lfloor\sqrt{p}\rfloor$. Having the performance of the optimal RF, it is compared to SVM and k-NN performances. As the Table 2 summarises, RF outperforms the other methods reaching a stable predictive accuracy of almost 78% ($ER_i = 0.22$) when m is fixed to $\lfloor\sqrt{p}\rfloor$. The performance of the other classifiers tends to change depending on the chosen seed, but their misclassification error rates are always greater than those obtained by RF.

RF with $m = \lfloor\sqrt{p}\rfloor$	SVM	k-NN
0.22	0.38	0.42

Table 2: Example of misclassification error rates

3.4 Results

The clustering and classification tasks yield contradictory results. The methods based on thematic or conceptual dimensions (i.e., topic modelling) show a mild dissimilarity, having a cosine of 0.82, between the two clusters. Moreover, the semantic dimension raises a further contradiction. Whereas the first words, *love* and *burn*, express opposite sentiments, according to different analyses reported in Table 1, the following items undermine this discrepancy due to their semantic neutrality. Indeed, several words are verbs (i.e., *want*), adverbs, or nouns whose meaning does not evoke any particular sentiment and, therefore, can not be identified as primers. In addition, the sentiment attributed to the term *burn* is also questionable due to the multiplicity of phrases and idioms within *burn* may be involved [13].

Interpreting the results of the classification task may also give rise to interesting conclusions. The better accuracy of the RF compared to the other classifiers has already been stated (Section 3.3), but its applicability in the real world is still questionable. Relevant insights may come from the variable importance plots (Fig.??), showing the importance of each predictor. Specifically, the Mean Decrease Accuracy plots the model accuracy losses when excluding each predictor, while the Mean Decrease Gini coefficient measures the contribution of each predictor to the node purity in the resulting RF [39]. The higher the value of mean decrease of accuracy or of Gini coefficient, the higher the importance of the predictor in the model. According to this dataset, the most important predictor is *danceability*, independently from the random seed set. As described in Section 3.1, *danceability* is computed by Spotify and combines different musical elements such as tempo, rhythm stability, beat strength, and overall regularity. Even though these audio features may indirectly support the hypothesis of the associative primer, the absence of a clear semantic difference between censured songs and not impedes any clear and meaningful insight.

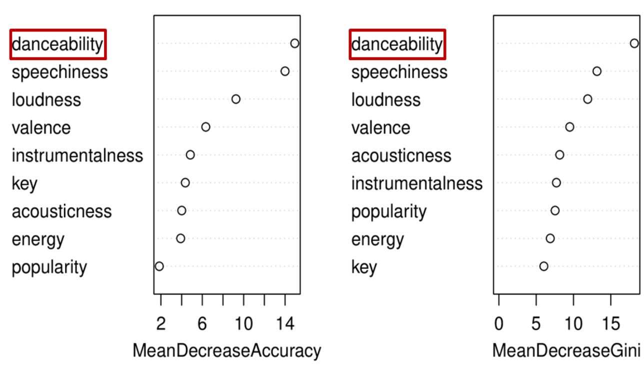

Figure 5: Variable Importance Plot of RF

4 Conclusion

The research was designed to detect hidden patterns within the songs censored after the 9/11 attacks and to formulate a valid explanation about their choice. Both

[13] https://www.merriam-webster.com/dictionary/burn
https://idioms.thefreedictionary.com/burning
https://idioms.thefreedictionary.com/burn+for

lyrics and audio features were involved in a multi-level model that performed a data clustering based on topics and a data classification based on audio features. Even though topic modelling and classification report moderate results, they raise some issues.

Firstly, the topic modelling and sentiment analysis, commonly employed in text mining, are performed on lyrics. This arises two problems: the absence of syntactical context and an ambiguous semantic. Thus, words have been clustered based on their co-occurrences within and between documents and then labelled according to predefined ontologies. This process can neither contextualize words nor capture their polysemy.

Secondly, the research question was focused on the possibility of generalizing the conclusions. However, the results emphasize a general applicability due to the scarcity of terms directly related to the attacks (e.g., airplane). Although the generality may seem a positive result, it should also suggest a problem in the general approach. Thus, the choices may be justified by reasons either unrelated to the hypothesis of associative primer (e.g., political reasons [10]), or not detectable by the multi-level model. According to the model, the analysis on lyrics and audio features is run on distinct levels involving any interaction between them. Interesting insights and a higher compliance with the real-world information processing mechanism may come from a multi–level model with integrated cross level interactions.

Another issue regards the classification task and its implementation which includes impure audio features. The assertion of impurity refers to Spotify's procedure which combines multiple musical features. A suitable example is *danceability*. As stated in Section 3.4, it is the most important predictor in the RF and combines four different musical features (Section 3.1), but any explanation regarding the relative impact of each feature is given. However, downloading audio features from Spotify had the main advantage of dealing with data already processed and bounded between 0 and 1. Thus, further studies dealing with pure audio features should face the additional challenge of data standardization.

Lastly, the scope of the research should be enlarged including other events that limited the diffusion of musical products, as the tsunami of 2004 and the banned song *"Die perkefte Welle - The perfect wave"*[14], or imposed controversial guidelines claiming to protect direct and indirect victims. Indeed, varying the context and the culture involved may lead to more significant of partially different conclusions.

References

[1] A. Cohen, "Music serves as a vehicle in multimedia contexts," in *Proceedings of the Joint Meeting of the Acoustical Society of America and the European Asoustical Society*, 1999.

[2] P. Evans and E. Schubert, "Relationships between expressed and felt emotions in music," *Musicae Scientiae*, vol. 12, no. 1, pp. 75–99, 2008.

[3] N. Steinbeis and S. Koelsch, "Affective priming effects of musical sounds on the processing of word meaning," *Journal of Cognitive Neuroscience*, vol. 23, no. 3, pp. 604–621, 2011.

[4] I. Peretz, M. Radeau, and M. Arguin, "Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs," *Memory & cognition*, vol. 32, no. 1, pp. 142–152, 2004.

[5] S. Johnson and A. Halpern, "Semantic priming of familiar songs," *Memory & cognition*, vol. 40, no. 4, pp. 579–593, 2012.

[6] P. Janata, S. Tomic, and S. Rakowski, "Characterisation of music-evoked autobiographical memories," *Memory*, vol. 15, no. 8, pp. 845–860, 2007.

[7] S. de l'Etoile, "The effectiveness of music therapy in group psychotherapy for adults with mental illness.," *The Arts in Psychotherapy*, 2002.

[8] S. Hecker, "Music for advertising effect," *Psychology & Marketing*, vol. 1, no. 3-4, pp. 3–8, 1984.

[9] N. Strauss, "The pop life; after the horror, radio stations pull some songs," *The New York Times*, 2001.

[10] M. Scherzinger, "Double voices of musical censorship after 9/11," in *Music in the post-9/11 world*, pp. 123–154, Routledge, 2013.

[11] M. Foucoult, "Discipline and punish," *A. Sheridan, Tr., Paris, FR, Gallimard*, 1975.

[12] S. Dubin, "Censorship: The knot that binds power and knowledge," 1991.

[13] L. Huddy, N. Khatib, and T. Capelos, "Trends: Reactions to the terrorist attacks of september 11, 2001," *The Public Opinion Quarterly*, vol. 66, no. 3, pp. 418–450, 2002.

[14] W. Schlenger, J. Caddell, and et al., "Psychological reactions to terrorist attacks: findings from the national study of americans' reactions to september 11," *Jama*, vol. 288, no. 5, pp. 581–588, 2002.

[15] T. Pyszczynski, S. Solomon, and J. Greenberg, "Terror management research: Coping with conscious and unconscious death-related thoughts.," in *In the wake of 9/11: The psychology of terror.*, ch. 3, pp. 37–70, American Psychological Association, 2003.

[14] https://en.wikipedia.org/wiki/Perfekte_Welle
The article published by the German newspaper *Der Spiegel*: https://www.tagesspiegel.de/gesellschaft/medien/die-perfekte-welle-kein-song-fuers-radio/573744.html

[16] G. Kvavilashvili, L.and Mandler, "Out of one's mind: A study of involuntary semantic memories," *Cognitive Psychology*, vol. 48, no. 1, pp. 47–94, 2004.

[17] M. Jungers, J. Hupp, and S. Dickerson, "Language priming by music and speech: Evidence of a shared processing mechanism," *Music Perception: An Interdisciplinary Journal*, vol. 34, no. 1, pp. 33–39, 2016.

[18] I. Peretz and M. Coltheart, "Modularity of music processing," *Nature neuroscience*, vol. 6, no. 7, pp. 688–691, 2003.

[19] C. Rogalsky, F. Rong, K. Saberi, and G. Hickok, "Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging," *Journal of Neuroscience*, vol. 31, no. 10, pp. 3843–3852, 2011.

[20] F. Rachman, R. Sarno, and C. Fatichah, "Music emotion classification based on lyrics-audio using corpus based emotion.," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 8, no. 3, 2018.

[21] A. Collins and E. Loftus, "A spreading-activation theory of semantic processing.," *Psychological review*, vol. 82, no. 6, p. 407, 1975.

[22] J. Mace, "Priming involuntary autobiographical memories," *Memory*, vol. 13, no. 8, pp. 874–884, 2005.

[23] Z. Peynírcíğlu, A. Tekcan, and *et al.*, "Name or hum that tune: Feeling of knowing for music," *Memory & cognition*, vol. 26, no. 6, pp. 1131–1137, 1998.

[24] S. Koelsch, E. Kasper, and *et al.*, "Music, language and meaning: brain signatures of semantic processing," *Nature neuroscience*, vol. 7, no. 3, pp. 302–307, 2004.

[25] R. Mayer, R. Neumayer, and A. Rauber, "Combination of audio and lyrics features for genre classification in digital audio collections," in *Proceedings of the 16th ACM international conference on Multimedia*, pp. 159–168, 2008.

[26] R. Mayer and A. Rauber, "Musical genre classification by ensembles of audio and lyrics features," in *Proceedings of International Conference on Music Information Retrieval*, pp. 675–680, 2011.

[27] K. Dakshina and R. Sridhar, "Lda based emotion recognition from lyrics," in *Advanced Computing, Networking and Informatics-Volume 1*, pp. 187–194, Springer, 2014.

[28] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[29] J. Pérez-Verdejo, C. Piña-García, M. Ojeda, A. Rivera-Lara, and L. Méndez-Morales, "The rhythm of mexico: an exploratory data analysis of spotify's top 50," *Journal of Computational Social Science*, vol. 4, pp. 147–161, 2021.

[30] K. Middlebrook and K. Sheik, "Song hit prediction: Predicting billboard hits using spotify data," *arXiv preprint arXiv:1908.08609*, 2019.

[31] J. Downie, "Music information retrieval," *Annual review of information science and technology*, vol. 37, no. 1, pp. 295–340, 2003.

[32] I. Kamehkhosh, G. Bonnin, and D. Jannach, "Effects of recommendations on the playlist creation behavior of users," *User Modeling and User-Adapted Interaction*, vol. 30, no. 2, pp. 285–322, 2020.

[33] S. Sangnark, M. Lertwatechakul, and C. Benjangkaprasert, "Thai music emotion recognition by linear regression," in *Proceedings of the 2018 2nd International Conference on Automation, Control and Robots*, pp. 62–66, 2018.

[34] G. Sharma and M. Murty, "Mining sentiments from songs using latent dirichlet allocation," in *International Symposium on Intelligent Data Analysis*, pp. 328–339, Springer, 2011.

[35] M. Omar, B. On, I. Lee, and G. Choi, "Lda topics: Representation and evaluation," *Journal of Information Science*, vol. 41, no. 5, pp. 662–675, 2015.

[36] R. Baeza-Yates, B. Ribeiro-Neto, and *et al.*, *Modern information retrieval*, vol. 463. ACM press New York, 1999.

[37] B. Larsen and C. Aone, "Fast and effective text mining using linear-time document clustering," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 16–22, 1999.

[38] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[39] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, vol. 112. Springer, 2013.

[40] J. Xin and B. Rongfang, "Random forest and pca for self-organizing maps based automatic music genre discrimination.," pp. 414–417, 01 2006.