Le Recommandeur: A Music Recommendation System Enhanced by NLP

Abdullah Palaz [1] **Mete Mert Birdal** [1] **Gökalp Özer** [1]

Abstract

Music has an important place in our daily life. According to research, people love music because sound is a part of human nature. We create a psychological bond between sounds and/or songs hence we feel closer to music. Most people listen to music at work, at the gym, or while relaxing at home. People have different tastes in music because music is a nearly limitless and differing field. When we feel energetic we listen to upbeat songs that makes us even more energetic and engaging, and when we are sad, we listen to slow, sad ones that we correlate with ourselves and the situation we are in. But sometimes, there is no song you know that you want to listen to because as humans, we tend to just get bored of what we have. These attitudes drive us to discover new songs all the time. This project's goal is to recommend you music that you would like and bring them to you. With audio features of the music you love, this model aims to recommend you the best music you have ever listened to.

1. Introduction

We listen to music all day and every day. Not just every day but everywhere, doing everything too. We work and listen to music, we are at the gym, working out, we listen to music, we study and we listen to music, we travel and we listen to music and this list can go way longer. But listening to that much music creates a problem where this particular person that listens to too much music becomes new musicless. So, it is an urge to find new music to listen to. This project's goal is to recommend you music that you would like and bring them to you. As known, music is just audio and, audio is feature-rich data. What we want is to create a music recommendation system that uses audio features and lyrics of the music to recommend songs that you might like. For recommendation, it is being planned to use Linear Regression and Logistic Regression. Usage of this model will be easy too. The user will simply rate some music and Le recommandeur will retrieve the ones that are similar to that data point.

[1]Department of Computer Engineering, Hacettepe University, Ankara, Turkey. Correspondence to: Abdullah Palaz <abdullahpalaz@hacettepe.edu.tr>, Mete Mert Birdal <metebirdal@hacettepe.edu.tr>, Gökalp Özer <gokalpozer@hacettepe.edu.tr>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

2. Related Work

There are two main approaches in recommendation systems, one of them is using features of music and the other is applying collaborative filtering and grouping music. Collaborative filtering is a method of making automatic predictions about the interests of a user by collecting information from users like him/her. Both these approaches are well developed, and many different machine learning models are being used in these recommendation systems such as Logistic Regression, KNN, and Deep Learning. There are real-life examples, other than academic research. Social platforms that include music like Spotify and Youtube, collect information from their users and recommend music for them. They use the listening activity of their users and bring similar music to them. Recent works on that matter are more focused on enhancing these models by using other attributes from different data sources. In 2018 there was a work enhancing music recommendation algorithms by using the "mood" of a person (extracted by Wearable Physiological Sensors). Another work from 2018 was trying to improve a music recommendation algorithm by finding emotion from that song's lyrics (this work was on Turkish songs) using Naive Bayes. In 2019, there was a project trying to extract emotion from the music's audio. In 2020, there was an NLP project trying to find the "type" of music based on their lyrics (This project was in Bengali.). The problem with enhancement using music lyrics is that the model will not be able to recommend songs in other languages and will not be able to recommend music that does not contain lyrics. Our model is not undiscovered but rather, it combines multiple aspects of music to recommend you new ones which is why we believe that it will be a successful and strong model.

3. Dataset

The dataset consists of about 2.2k music with 19 genres. This genre number will change since Spotify (where the mu-

sic tags are retrieved) does not hold this genre information that general, instead, there are Australian pop, Canadian pop, dance-pop, etc., where there should only be pop. The lyrics of the music are still being collected, and this is purely done by hand.

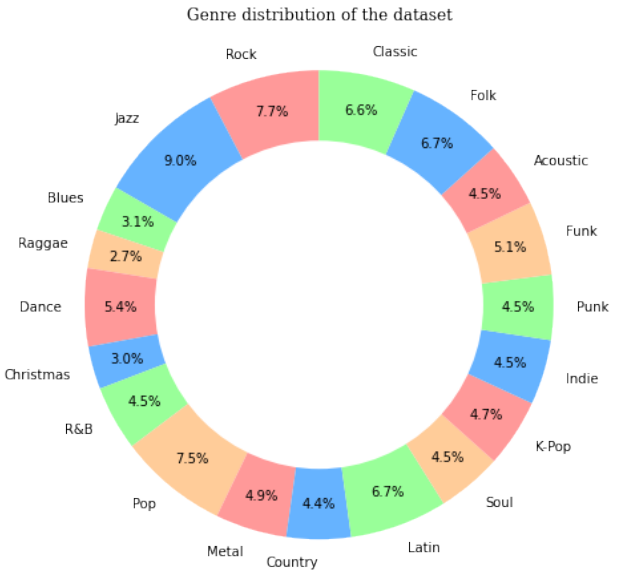

Figure 1. Music genres and their distribution in the dataset.

3.1. Audio Features

Audio features are the description of an audio signal. There are 9 audio features that we are using where there are 30 statistical variables that define these features. All features represent unique and useful information. These features are:

- Chroma
- MFCCs
- Spectral Centroid
- Spectral Contrast
- Spectral Roll Off
- Zero Crossing Rate
- Beat
- Tempogram
- Melspectogram

3.2. Lyrics Feature

The model will be enhanced using the lyrics. We will extract the mood of the music using its lyrics. In this way, we believe that we will improve the recommendation capacity of our model. The problem with this approach is that now our model can only recommend English songs, and it is still ambiguous if it can recommend music with no lyrics.

3.3. Representation of a music

These representations are in the very base format. All other features of the audio are extracted from the array that has been used to create these plots. We visualized music in waveplot and spectrogram forms. The visual display shows various shapes that change the size, color, and brightness in correlation with the music.

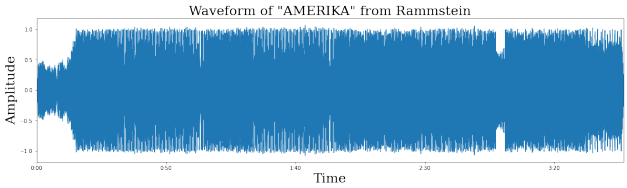

Figure 2. Waveplot form of a song.

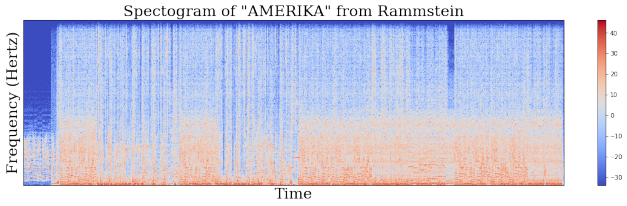

Figure 3. Spectogram form of a song.

4. Methodology

We want to recommend songs using their features in Section 3.1. Initially what we wanted was to use 1D CNN to get a representation of the song, then extract some features of the song using librosa and extract the mood of the song using its lyrics. (See Figure 4.).Then, we wanted to concatenate these features to create a whole representation of a song and feed this to a machine learning model.

But due to time and computational complexities, we decided that this is not feasible to do. So, we removed the 1D CNN part for audio representation, and we achieved the approach demonstrated in Figure 5.

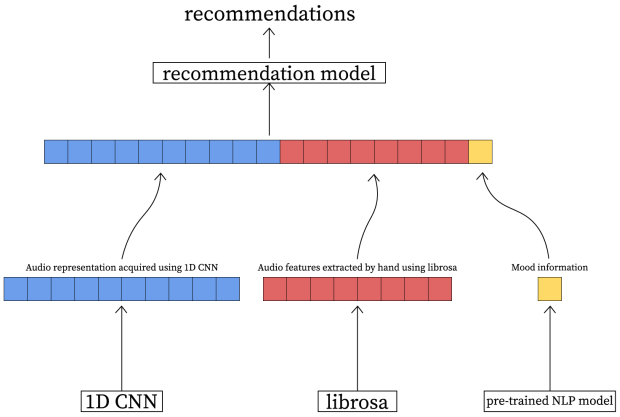

Figure 4. Figure demonstrating our initial approach.

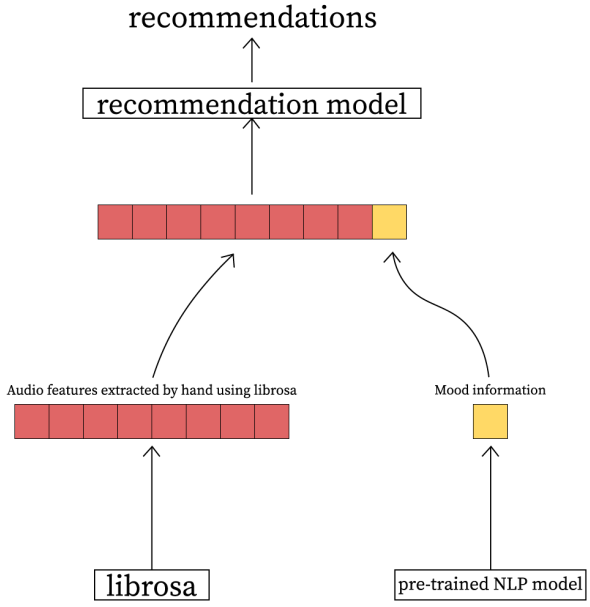

Figure 5. Figure demonstrating our final approach.

5. Discussions and Future Processes

Librosa library has been used to extract audio features. After that, we have a few models on our minds as the recommendation model. Before talking about the model that we consider using, it is needed to discuss the rating approach we have in our minds. Rating will either be 0 - 10 or binary, as 0 (or -1) and 1. Rating from 0 to 10 may yield better results but when we consider this model as a real-life application model, no one will just rate music just to get new recommendations. Especially, in a music app like Spotify, the binary rating can be used easily because the user does not have to rate songs, since there is already a ”Liked Songs” section where the rating of this user can be taken from. When we address this problem, we will select a recommendation model. This model is thought to be a classifier but as mentioned, it is still not decided. We are considiring CNN, ANN, Linear Regression, Logistic Regression right now. This considiration list may expand, since we are about to implement this part of the project very soon.

Accuracy will be measured using the user one more time. After the user got the recommendations, the user will rate the recommendations. Since all the recommended songs should be rated ”1” in binary rating and about ”7” or higher in 0 to 10 rating, these ratings will be ground truths and if the user will dislike a recommended song, then accuracy will decrease automatically. Since the accuracy will base on the user, it will not be something that represents the truth. We considered this approach because a user can’t rate all the songs in our dataset, and if one day the dataset expands, where will the rating end?

References

Coban, O. and Ozyer, G. T. Music genre classification from turkish lyrics. In *2016 24th Signal Processing and Communication Application Conference (SIU)*, pp. 101–104, 2016. doi: 10.1109/SIU.2016.7495686.

Li, T., Chan, A., and Chun, A. Automatic musical pattern feature extraction using convolutional neural network. *Lecture Notes in Engineering and Computer Science*, 2180, 03 2010.

Nath, D., Roy, A., Shaw, S. K., Ghorai, A., and Phani, S. Textual lyrics based emotion analysis of bengali songs. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 39–44, 2020. doi: 10.1109/ICDMW51313.2020.00015.

Toiviainen, P. and Eerola, T. Method for comparative analysis of folk music based on musical feature extraction and neural networks. 09 2001.