MOODLIST - PLAYLISTS TUNED TO YOUR MOOD

A MOOD CLASSIFIER BASED ON RECENT MUSIC STREAMING HISTORY

Harshit Sohaney
Student# 1006982482
harshit.sohaney@mail.utoronto.ca

Maisey Perelonia
Student# 1006681440
maisey.perelonia@mail.utoronto.ca

Anne Chow
Student# 1007332593
anne.chow@mail.utoronto.ca

Abeer Fatima
Student# 1006845416
ab.fatima@mail.utoronto.ca

ABSTRACT

Moodlist has the power to determine a user's mood based on the songs they recently listened to. This is a well-established concept, given that multiple studies have been conducted on Music Emotion Recognition (MER) models where researchers used multi-class classification algorithms to detect moods. Based on the scope of Moodlist, the team has gathered appropriate datasets to train and test our model. The project will utilise a Convolutional Recurrent Neural Network (CRNN) to synergise the best features of CNN and RNN models, allowing for feature classification and sequence prediction. Several MER-related ethical issues are considered, specifically, data privacy and mining, emotional regulation, and geographic biases. Furthermore, the team has laid out a detailed timeline with internal and external milestones for its project. There are several group risks as well as technical risks, for which the team has outlined preventative measures.

—-Total Pages: 9

1 INTRODUCTION

Music streaming took over the world in 2010, replacing traditional forms of listening such as radio and vinyl in the process. This new phenomenon of being able to discover new music while leaving a digital footprint of listening habits introduced an era of music recommendation into the world.

Deep learning has made it possible to discover new music daily in the form of playlists curated for a listener. For example, the streaming giant Spotify has its ”discover weekly” playlist to thank for its industry success (spo (2020)).

Our project involves detecting a listener's mood given a set of songs they listened to recently. As a large portion of our team has a musical background, our motivations for the project were influenced by our musical expertise. For listeners, we want to be able to cater to how they might be feeling and recommend songs appropriately. Moreover, this offers music creators with the opportunity to be discovered based on both a listener's taste and their mood.

Our team's goal is to have Moodlist help listeners find music based on both how they are feeling and their overall music taste. We believe that deep learning can help achieve this by learning a user's mood and using it as a parameter in other recommendation systems. The following sections will layout our plans to execute the project.

2 ILLUSTRATION

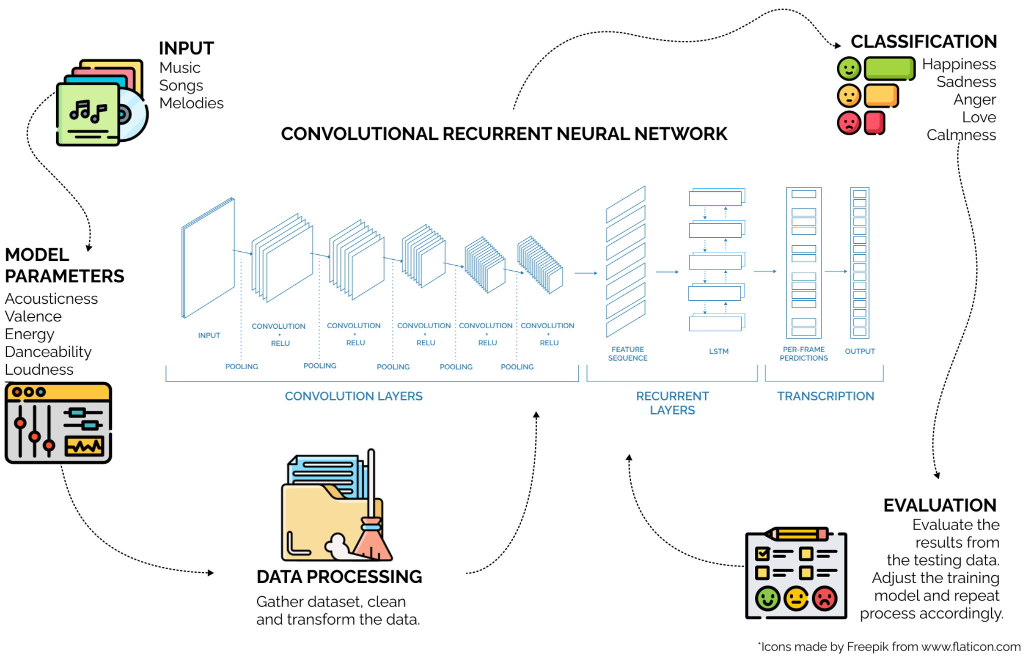

Figure 1: Project Overview

3 BACKGROUND & RELATED WORK

As recommendation systems become increasingly personalized, the field of Music Emotion Recognition (MER) grows rapidly as a research area. The ability to recognize a user's mood based on what they have been listening to can help create engagement on platforms with more personalized recommendations. Some recent studies in this field are described below.

3.1 MOOD DETECTION BASED ON LAST SONG (FERDIANA ET AL. (2022))

This study uses multi-class classification algorithms to detect a person's mood based on the last song they listened to. The dataset consists of both lyrics and songs to encapsulate both aspects of music. They found success in dividing moods into a Four-Dimensional Mood Scale (FDMS) which measures mood based on positive energy, tiredness, negative activation, and relaxation.

After splitting their Spotify and Genius testing data into 80% training and 20% testing data, the results of the experiment yielded a 91% accuracy using the FastTreeOva algorithm.

3.2 MUSIC MOOD CLASSIFICATION SYSTEM FOR STREAMING PLATFORM ANALYSIS (CHEN ET AL. (2021))

The study presented at an IEEE conference in Taiwan 2021, highlights the use of Convolutional Neural Networks (CNN) on music sonograms. Chen et al. used a Forest Tree Classifier and other algorithms to bridge both songs and lyrics into their DL model. They conducted feature extraction on both datasets and then concatenated them together. This concatenated feature set was then passed into their fully connected models to examine results. We plan on concatenating our datasets in a similar manner. Figure 2 lays out the architecture of their study. They achieved an 84.52% accuracy on a testing set of 193 songs.

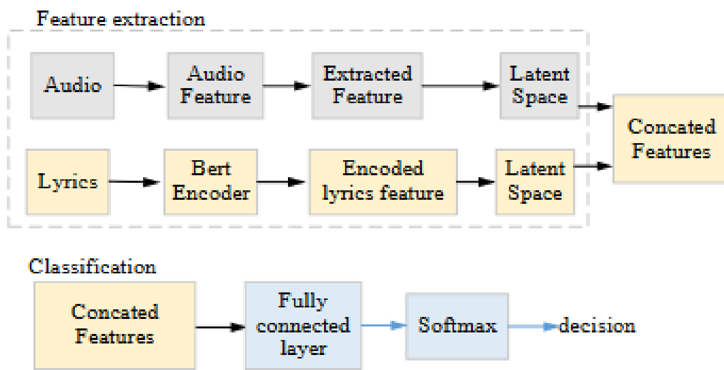

Figure 2: Chen et al. study architecture

Having established a background for our project, the following section discusses how the data for our model shall be processed.

4 Data Processing

Data Processing can be broken down into 5 stages (Duggal (2022)):

1. Data Collection
2. Preparation
3. Input & Processing
4. Output & Interpretation
5. Storage

4.1 Data Collection

At this stage, the collection of raw data is key to building the foundations of data processing. To ensure the results from the data are accurate, the source and data itself must be precise and valid. As shown in Table 1, we will be using the ”Spotify Million Playlists Dataset” (Spotify (2018)) as our training data along with the Spotify API (Spotify) to classify moods based on the song. In addition, we will be utilizing the MER dataset, MIREX (Panda et al. (2013)), as our testing data.

Dataset	Contents	Year Published	Size
Multi-modal MIREX Emotion	903 audio clips - 30 seconds each	2013	320 MB
Spotify Million Playlist Dataset	2 million tracks	2018	5 GB

Table 1: Datasets for our model.

4.2 PREPARATION

Before the data can be used as input, it must be cleaned to guarantee validity, accuracy, completeness, consistency, and uniformity. The data cleaning workflow can be described in a 6 step process (tableau):

1. Duplicate & Irrelevant Data
2. Structural Errors
3. Outliers
4. Missing Data
5. Quality Assessment & Validation
6. Storage

Duplicate & Irrelevant Data Collecting datasets from multiple sources can result in the aggregation of duplicate or irrelevant data. To ensure that the dataset is reflective of our project purpose, we must remove these unwanted data points.

Structural Errors Any inconsistencies on the naming, labeling, and classification of the data must be fixed to garner a coherent uniform dataset.

Outliers Outliers should be checked to see if they are relevant to the purpose of the data. Before removing an outlier, we must ensure that the outlier itself is invalid.

Missing Data We must handle any data with missing values to ensure our algorithm will not break. To do so, we can remove those pieces of data, fill in any missing values based on patterns found in our dataset, or nullify any missing values.

Quality Assurance & Validation After cleaning the data, we should review the dataset to reaffirm its logic and ensure it aligns with the scope of our project.

4.3 INPUT & PROCESSING

To generate a uniform dataset, we will trim each audio excerpt to 30 seconds from a uniformly distributed starting point and make sure that all samples are mutually exclusive. From there, we will convert the audio files to .CSV files so that the files can be inputted to our training model.

4.4 OUTPUT & INTERPRETATION

Similar to the input, the output will be in a .CSV format with corresponding mood tags on each song. We will use our model on the testing dataset, MIREX, and compare our results with the emotion tags within the MIREX data. Based off the outcome of the comparison, we will adjust the model's hyper-parameters, and reiterate through the process accordingly.

4.5 STORAGE

We will store all the resulting data away in our GitHub for future reference. This allows for the team to quickly retrieve any information/data if ever needed.

The next section will explore further into the architecture of our model.

5 Architecture

Both Convolutional and Recurrent Neural Networks are known to be successful at classification problems. Traditionally, CNNs have been used for image classification and RNNs for sequence prediction (Brownlee (2019)). For the purposes of this project, we want to be able to classify a user's mood by taking their listening history into account. This is why we have chosen to make a combined Convolutional Recurrent Neural Network (CRNN), where the CNN handles feature classification and the RNN manages sequence prediction. A commonly used RNN network in practice is the Long Short Term Memory (LSTM) network which is capable of learning long-term dependencies. Figure 3 highlights this architecture in a preliminary design.

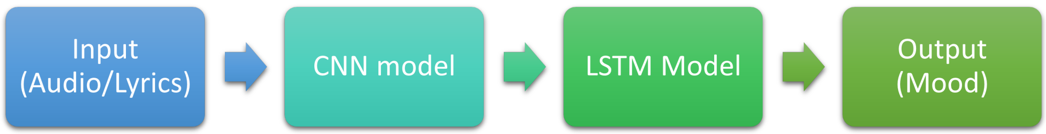

Figure 3: CRNN Preliminary Architecture

6 Baseline Model

Our baseline model is a support vector machine (SVM). A SVM model maps linearly and non-linearly separable data onto a higher dimensional space in order to categorize the data points. This is done by finding a separator between the data categories and then drawing it as a hyperplane (e.g. a plane in 3D and a line in 2D).

We chose this model as a baseline because its calculations are relatively simple and its results are easy to interpret. SVM uses binary classification, so when testing, multiple layers will be necessary as we will be classifying songs into more than two moods. In order to plot each data point onto the graph, we will use different song characteristics like tempo, valence, and loudness. At each layer, the two groups of items should then be separable using a hyperplane. The result would be an easily reproducible baseline model which can be used as a benchmark for our final model. In the past, accuracy rates up to 76.6% have been achieved (Mutiara et al. (2016)), so our benchmark for a successful project means exceeding this accuracy, past at least 77%.

The next section considers ethical repercussions that we have to consider while we are working on this project.

7 Ethical Considerations

In today's day and age, most people use music streaming platforms like Spotify to listen to music and assume that any personal data given to these platforms is simply for identification purposes. In reality, however, an individual's cookie data and IP address can be used to track them and retrieve potentially sensitive information on their devices. This processing of large quantities of data to draw conclusions and thus create new data is known as data mining and is a violation of freedom from surveillance (Aloshyas (2020)).

On a similar note, an individual's mental privacy is compromised as MER allows for the detection of their mental state which they may not wish to share. Seeing their results may, furthermore, put social pressure on some to self-regulate their emotions to become what is considered 'normal' (Steinert & Friedrich (2019)). An inaccurate model may adversely impact users' mental health by causing self-doubt and, in extreme cases, identity crises.

On a larger scale, deep learning models are inevitably biased and raise several ethical questions. The effectiveness of such a preliminary model may be limited to music from a given region in the world, namely those who consider English one of their official languages. Due to an uneven distribution in data, it may even associate certain moods with specific languages. As a result, this can create discriminatory bias and exclude individuals of nations with diverse languages and sounds from benefiting from this model.

The following section discusses the project plan.

8 Project Plan

The figures below display our team's Notion database which consists of all the higher-level (Figure 4) and lower-level (Figure 5) deadlines pertaining to our project with assignee(s), current status (at time of writing), priority, deadline, and label(s).

Name	Status	Assignee	Deadline	Priority	Label(s)
Project Ideation	Completed	Abeer Fatima, Harshit Sohaney, Maisey Perelonia, Anne Chow	October 9, 2022	High 🔥	Milestone
Data Collection	In Progress	Anne Chow	October 13, 2022	High 🔥	Milestone
SUBMIT Project Proposal	Next Up	Abeer Fatima, Harshit Sohaney, Maisey Perelonia, Anne Chow	October 13, 2022	High 🔥	Milestone
Project Proposal	In Progress	Abeer Fatima, Harshit Sohaney, Maisey Perelonia, Anne Chow	October 14, 2022	High 🔥	Course Deadline
Data Preparation (Cleaning)	Next Up	Harshit Sohaney	October 30, 2022	High 🔥	Milestone
Input [usable data]	Next Up	Abeer Fatima	November 6, 2022	Medium	Milestone
Data Processing	Not Started	Maisey Perelonia, Abeer Fatima, Anne Chow, Harshit Sohaney	November 20, 2022	Medium	Milestone
Model Evaluation/Interpretation	Not Started	Maisey Perelonia, Abeer Fatima, Anne Chow, Harshit Sohaney	November 24, 2022	Medium	Milestone
Project Presentation	Not Started	Abeer Fatima, Harshit Sohaney, Maisey Perelonia, Anne Chow	November 25, 2022	Low	Course Deadline
Progress Report	Not Started	Abeer Fatima, Harshit Sohaney, Maisey Perelonia, Anne Chow	December 2, 2022	Low	Course Deadline

Figure 4: Higher-Level Overview of Team Progress

Name	Status	Assignee	Deadline	Priority	Label(s)
Project Ideation	Completed	Abeer Fatima	October 9, 2022	High	Milestone
Team Meeting	Completed	Abeer Fatima	October 9, 2022	High	Meeting
[first draft] Illustration, Data Procession	Completed	Anne Chow	October 11, 2022	High	Internal Deadline
[first draft] Baseline Model, Risk Registers	Completed	Maisey Perelonia	October 11, 2022	High	Internal Deadline
[first draft] Ethical Considerations + Project Plan	Completed	Abeer Fatima	October 11, 2022	High	Internal Deadline
[first draft] Introduction + Background & Related Work	Completed	Harshit Sohaney	October 11, 2022	High	Internal Deadline
Data Collection	In Progress	Anne Chow	October 13, 2022	High	Milestone
SUBMIT Project Proposal	Next Up	Abeer Fatima	October 13, 2022	High	Milestone
Background and Related Work	In Progress	Harshit Sohaney	October 13, 2022	High	Internal Deadline
Introduction	In Progress	Harshit Sohaney	October 13, 2022	High	Internal Deadline
Risk Register	In Progress	Maisey Perelonia	October 13, 2022	High	Internal Deadline
Baseline Model	In Progress	Maisey Perelonia	October 13, 2022	High	Internal Deadline
Project Plan	In Progress	Abeer Fatima	October 13, 2022	High	Internal Deadline
Ethical Considerations	In Progress	Abeer Fatima	October 13, 2022	High	Internal Deadline
Architecture	Completed	Abeer Fatima	October 13, 2022	High	Internal Deadline
Data Processing OPEN	In Progress	Anne Chow	October 13, 2022	High	Internal Deadline
Illustration	In Progress	Anne Chow	October 13, 2022	High	Internal Deadline
Team Meeting + Worksession	Completed	Abeer Fatima	October 16, 2022	Low	Meeting
Team Meeting + Worksession	Completed	Abeer Fatima	October 23, 2022	Low	Meeting
Team Meeting + Worksession	Completed	Abeer Fatima	October 30, 2022	Low	Meeting
Data Preparation (Cleaning)	Next Up	Harshit Sohaney	October 30, 2022	High	Milestone
Team Meeting + Worksession	Completed	Abeer Fatima	November 6, 2022	Low	Meeting
Input [usable data]	Next Up	Abeer Fatima	November 6, 2022	Medium	Milestone
Team Meeting + Worksession	Completed	Abeer Fatima	November 13, 202	Low	Meeting
Team Meeting + Worksession	Completed	Abeer Fatima	November 20, 202	Low	Meeting
Data Processing	Not Started	Maisey Perelonia	November 20, 202	Medium	Milestone
Model Evaluation/Interpretation	Not Started	Maisey Perelonia	November 24, 202	Medium	Milestone
Team Meeting + Worksession	Completed	Abeer Fatima	November 27, 202	Low	Meeting

Figure 5: Lower-Level Overview of Team Progress

Evidently, stringent project management measures have been taken to ensure our team is consistently on task and transparent with regards to progress. Tasks have and will continue to be divided according to individual skill-sets, interests, and workloads. The team will maintain communication via WhatsApp to efficiently resolve minor issues and will meet every Sunday 10AM-11AM for weekly updates and 11AM-1PM for a synchronous work session. We will write our code on GitHub and create branches, pull requests, and formatting rules to avoid overwriting each other's code. All in all, clear, regular communication and equitable distribution of tasks will be our priority.

On a related note, the next section discusses the team's risk management strategies.

9 Risk Registers

There are multiple risks that may occur, both technical and nontechnical, especially in a team where one person's actions can affect everybody else. One risk that may occur during the course of this project is having a teammate drop out of the course, voluntarily or involuntarily. In this case, we would move forward, reconvene, and split the leftover tasks between the remaining members.

Another risk involves not planning out our dates and deadlines appropriately, causing us to fall behind, or pushing back internal deadlines because we are unable to meet them. As a team, we need to hold ourselves accountable and enforce strict deadlines. If one of us cannot meet a deadline, then they will have to let the rest of the team know at least two days in advance. Additionally, we need to plan dates with a buffer in mind so that we are not rushing to finish things in case the tasks take longer than expected. If we do not meet the internal deadline, we will schedule an emergency meeting as soon as possible to finish up and redistribute any incomplete tasks.

One technical risk we may experience is overfitting the data in our training sets, resulting in a high testing data loss. This also introduces bias towards certain types of data such that it cannot properly classify data of other types. In order to mitigate this issue, we would have to adjust our hyperparameters so that our model is exposed just enough to our training data that it will learn enough but not memorize. This includes the number of layers, number of epochs, learning rate, as well as many iterations in order to see what hyperparameters are allowing the model to perform best.

Finally, our team must accept that it is not guaranteed that we will accomplish our end objective within the given time frame. In order to prevent this possibility to the best of our abilities, we have set milestone checkpoints throughout the semester to see if we are making progress and if our model is working.

10 GitHub Link

`https://github.com/maiseyperelonia/moodlist`

References

Spotify users have spent over 2.3 billion hours streaming discover weekly playlists since 2015. *Spotify*, Jul 2020.

Riyo Aloshyas. Spotify: Data privacy case study, Sep 2020. URL https://riyoalo.medium.com/spotify-data-privacy-case-study-a9167405c22d.

Jason Brownlee. Cnn long short-term memory networks, Aug 2019. URL https://machinelearningmastery.com/cnn-long-short-term-memory-networks/.

Yu-Chia Chen, Zih-Ching Chen, and Chih-Hsien Hsia. Music mood classification system for streaming platform analysis via deep learning based feature extraction. *2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, Sep 2021. doi: 10.1109/icce-tw52618.2021.9603205.

Nikita Duggal. What is data processing system? definition, cycle, types amp; methods [updated], Sep 2022. URL https://www.simplilearn.com/what-is-data-processing-article.

Ridi Ferdiana, Wiiliam Fajar Dicka, and Faturahman Yudanto. Mood detection based on last song listened on spotify, Sep 2022. URL https://journals.utm.my/aej/article/view/16834.

Achmad Benny Mutiara, Rina Refianti, and Nadia R.A. Mukarromah. Musical genre classification using svm and audio features. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 14(3):1024, 2016. doi: 10.12928/telkomnika.v14i3.3281.

R Panda, R Malheiro, B Rocha, A Oliveira, and R P Paiva. Multi-modal music emotion recognition: A new dataset, methodology and comparative analysis. 2013.

Spotify. Web api. URL https://developer.spotify.com/documentation/web-api/.

Spotify. Spotify million playlist dataset challenge: Challenges, 2018. URL https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge#dataset.

Steffen Steinert and Orsolya Friedrich. Wired emotions: Ethical issues of affective brain–computer interfaces. *Science and Engineering Ethics*, 26(1):351–367, 2019. doi: 10.1007/s11948-019-00087-2.

tableau. Guide to data cleaning: Definition, benefits, components, and how to clean your data. URL https://www.tableau.com/learn/articles/what-is-data-cleaning.