Generating a Mood-centric Spotify Playlist using Audio Feature Classification and Sentiment Analysis

Keshan Moosai
Department of Computing and Information Technology
University of the West Indies
St. Augustine, Trinidad

Vinaayaka Seelal
Department of Computing and Information Technology
University of the West Indies
St. Augustine, Trinidad

Virendra Narine
Department of Computing and Information Technology
University of the West Indies
St. Augustine, Trinidad

Shaniah Baldeo
Department of Computing and Information Technology
University of the West Indies
St. Augustine, Trinidad

Abstract— This paper investigates an application that automatically generates playlists tailored to a user's emotional state. The application leverages two machine learning models working in tandem. The first, a Music Mood Classifier (MMC), analyses audio features like tempo, timbre, key, and energy to classify songs into eight distinct moods (Happy, Exuberant, Energetic, Frantic, Sad, Depression, Calm, Contentment). The second model, an Emotion Detection Model (EDM), focuses on the user. It analyses user input text to identify the underlying emotional state and maps it to one of the eight mood categories. An LSTM (Long Short-Term Memory) network is employed within the EDM for superior text-based emotion detection. The system was evaluated using precision and accuracy metrics, achieving a precision of 0.67 for the LSTM model. Peer evaluations also indicated the effectiveness of the playlists in representing the intended mood. This paper acknowledges limitations such as dataset size and subjectivity of emotions. Future work includes data augmentation and incorporating sentiment analysis from song lyrics to enhance the model's accuracy. This research contributes to the field of music mood classification and emotion detection from text, with potential applications in music recommendation systems and personalised playlist generation.

Keywords— Music Mood Classifier, Emotion Detection Model, Sentiment Analysis, Recurrent Neural Network, Logistic Regression, Long Short-Term Memory

1. INTRODUCTION

1.1 Motivation

The Caribbean region is a cultural hub with deep rooted history in music and expression. Music helps people express their sentiment and emotions [8] Music helps tie the Caribbean region to the rest of the world. Caribbean artists such as Bob Marly have captured the hearts of the world through Reggae. Modern-day musicians continue to captivate foreign audiences through the medium of Soca music during Carnival. Music is an important tool that can influence the state of mind, emotions and moods of both individuals and the environment that music is played in. Music also provides entertainment for individuals as well as being a therapeutic tool for one's emotions. Curating a musical selection has traditionally been a manual process as individuals would listen and choose songs that meet some emotional requirements.

1.2 Outline of Research

Automating this process requires extra processing on songs. Mood is not a metric that modern music platforms, such as Spotify, classify for individuals. However, breaking a song down into quantifiable musical components such as rhythm, harmony, and timbre can allow for the matching of songs to specific categories based upon expected data for each type of mood. These features were extracted from acoustical analysis from a sample of 44100 songs [1]. With this approach, our Music Mood Classifier (MMC) uses audio features extracted by the Spotify API to class songs into one of eight mood clusters. Supervised learning will be implemented to continuously class new songs added to the metadata set.

While the MMC deals with classing the moods of the songs, our Emotion Detection Model (EDM) deals with mapping an input text, provided by the individual, to one of eight mood clusters. Text analysis with supervised learning will be used to train the EDM. These two models create the link between the mood of the individual and the mood of the songs.

The underlying objective of this paper is to design an accurate algorithm that would yield a playlist of songs in conformance with a user's emotional state determined by the input phrase/sentence.

2. LITERATURE REVIEW

This literature review is structured into five essential sections, each contributing to our understanding of the intricate relationship between music, mood, and emotion. We commence by surveying related works, providing context for subsequent discussions. Next, we define mood and emotion, establishing terminological clarity. Subsequently, we delve into the classification of music into moods, exploring methodologies and frameworks. Following this, we examine emotion detection using sentiment analysis, uncovering insights into textual data analysis. Finally, we conclude by summarising key findings from the preceding sections and identifying avenues. Through these sections, we aim to provide a comprehensive understanding of the complex interplay between music, mood, and emotion, offering insights into this dynamic field.

2.1 Related Work

The investigation into the intricate relationship between music, mood, and emotion spans various research endeavours, each offering unique insights and methodologies. One avenue involves leveraging digital signal processing (DSP) techniques and music theory principles to analyse music and classify mood based on rhythmic, harmonic, and spectral features [1]. Simultaneously, efforts are underway to enhance music recommendation systems through novel visualisations and mood control mechanisms, facilitating users' exploration of new music genres [3]. Advancements in text-based mood detection have enabled real-time categorization of streaming text data into emotional classes, offering valuable insights into mood fluctuations [5]. Leveraging machine learning techniques, researchers aim to predict mood from textual data, with implications for sentiment analysis [5]. Furthermore, the development of innovative models, such as logistic regression [6] and convolutional long short-term memory (CLDNN) architectures [12], demonstrates the potential of deep learning methodologies to enhance music emotion recognition systems, achieving high accuracy in emotion classification. Collectively, these research pursuits underscore the multidisciplinary nature of studying music, mood, and emotion, contributing to our understanding of human emotional responses to music and paving the way for innovative applications in this dynamic field.

2.2 Definitions and Terminologies

Emotions and moods are fundamental aspects of human experience, influencing perception, cognition, and behaviour. While definitions may vary across disciplines, encompassing perspectives from psychology, neuroscience, sociology, philosophy, anthropology, and biology, overarching themes provide insights into their nature and manifestations.

2.2.1. Emotions

Emotions denote intense, transient states of feeling often triggered by specific stimuli or events [8]. They encompass a broad spectrum of affective experiences, including arousal, pleasure, and displeasure, and are characterised by complex interactions between subjective and objective factors mediated by neural and hormonal systems. Emotions give rise to affective experiences, cognitive processes, physiological responses, and expressive behaviours, contributing to adaptive responses to environmental stimuli.

2.2.2. Moods

Moods are enduring, less intense states of feeling that tend to lack a specific contextual stimulus. Unlike emotions, which are discrete and short-lived [8], moods persist over time and influence the overall emotional tone or disposition of an individual. They encompass a broad range of subjective experiences, shaping cognition, perception, and behaviour. Moods are characterised by their longer duration and lower intensity compared to emotions, often reflecting a person's prevailing emotional state [8].

2.2.3. Mood Classification vs. Emotion Classification

In the context of music analysis and playlist generation, distinguishing between mood and emotion is essential. Emotion classification focuses on identifying discrete affective responses to specific musical elements or events, capturing intense and short-lived emotional experiences. In contrast, mood classification emphasises the broader emotional atmosphere or tone conveyed by music, capturing enduring and pervasive affective states that shape the overall listening experience.

2.2.4. Audio Features

Audio features such as tempo, timbre, key, and energy play crucial roles in shaping the emotional and affective impact of music. Tempo influences the perceived pace and energy level of music, while timbre contributes to its unique sonic qualities and emotional resonance. Key, which was representational of the pitch, was deterministic of the sound frequency that was influential on emotional responses. Energy reflected the intensity and vigour of the music, shaping its emotional impact and mood-inducing potential via its loudness [1].

2.3 Classifying Music into Moods

Within the sphere of music mood classification, a prevailing methodology draws upon the insights of psychologist Robert Thayer. Thayer's model, which centres on the dimensions of energy and stress, partitions mood along a spectrum ranging from happy to sad and calm to energetic [1]. This framework delineates eight distinct mood categories, encompassing both the extremes of each dimension and their intersections, such as lively-calm or sorrowful-energetic.

In the context of Nuzzolo's research, an insightful exploration unveils a noteworthy investigation conducted by engineers hailing from the BNM Institute of Technology in Bangalore, India. This study embarked on an endeavour to evaluate intensity, timbre, pitch, and rhythm levels across a diverse selection of songs, spanning various mood states. Utilising a tailored algorithm, the researchers sought to distil meaningful audio features conducive to classification [1]. The classification process entailed a meticulous comparison of these extracted features with predefined threshold values corresponding to each distinct mood. Notably, Table 2 from Nuzzolo’s work encapsulates the average values of each audio feature identified across specific mood classifications, as gleaned from this experimental inquiry [1].

2.4 Emotion Detection using Sentiment Analysis

In the exploration of emotion detection through sentiment analysis, various methodologies have surfaced as robust options, each employing distinct techniques to discern and categorise emotional nuances within textual data. Notably, logistic regression models with TF-IDF encoding, TF-IDF combined with n-grams, and LSTM (Long Short-Term Memory) models have gained prominence for their efficacy in this field.

Logistic regression models with TFIDF encoding establish a robust framework for sentiment analysis by quantifying the relevance of each term in a document while mitigating the impact of commonly occurring terms across the corpus [11]. This methodology facilitates the identification of significant features indicative of different emotional states, thereby enabling accurate classification.

Building upon TF-IDF, the incorporation of n-grams enhances the granularity of sentiment analysis by capturing contextual information and syntactic structures within the text [15]. By considering sequences of adjacent words, TF-IDF combined with n-grams empowers the model to discern subtle nuances and complex patterns in emotional expression, thus refining the accuracy of sentiment classification.

Similarly, LSTM models, a type of recurrent neural network (RNN), have emerged as powerful tools for sentiment analysis, especially in scenarios involving sequential data like text [10]. The capability of LSTM to retain long-range dependencies and capture temporal dynamics within the text enables it to discern intricate emotional nuances and contextual cues, ultimately yielding nuanced and contextually informed sentiment classifications.

These methodologies offer distinct advantages in terms of feature representation, contextual understanding, and modelling capabilities, catering to diverse research needs and objectives within the field of emotion analysis. Consequently, they prove to be viable options for sentiment analysis in research endeavours, particularly in the context of emotion detection from user input regarding their mood in response to music.

2.5 Conclusion

Our exhaustive exploration of music, mood, and emotion literature yielded key insights essential for advancing research in this domain. We observed the nuanced definitions of "mood" and "emotion" and the effectiveness of both categorical and dimensional frameworks in capturing their essence. Notably, our review highlighted the superiority of LSTM models in emotion detection, particularly in discerning subtle emotional nuances within textual data. Leveraging this finding, we plan to integrate LSTM models into our research methodology to enhance the accuracy and depth of our analysis. By embracing these insights, we endeavour to understand the dynamic intersection of music, mood, and emotion.

3. METHODOLOGY

3.1 Data Collection and Preprocessing

The project utilised two primary datasets: the Google Emotions dataset and a dataset obtained from Kaggle containing 1.2 million data points. The Google Emotions dataset, comprising over 200 thousand records and featuring 28 columns, was particularly notable for its inclusion of 27 unique emotions. To facilitate analysis, this dataset underwent one-hot encoding to associate textual data with specific emotions. Additionally, the Spotify Track Features Dataset, containing 1.2 million records and 24 columns, was employed. From this dataset, only four columns were retained for analysis: 'id', 'energy', 'key', and 'tempo'. Furthermore, a new column named 'timbre' was introduced. The research referenced in the literature utilised features such as 'energy', 'key', 'tempo', and 'timbre' to classify the mood of songs into one of the eight emotions previously mentioned. Notably, 'timbre' values for each song were obtained through the Spotify Web API. These values, represented by 12-element vectors, underwent transformation into their respective linearly independent solutions, resulting in a singular value. The general formula is given:

$$\sum_{i=1}^{12} a_i \cdot u_i$$

$$a_i = 1$$

$$a_i \text{ represents the scalar}$$

$$u_i \text{ represents the basis}$$

Whilst the key and timbre features were normalised to values between 0 and 1, the energy feature was already provided to us in a normalised form. The tempo column was not normalised due to the absence of official documentation or research providing a maximum tempo value. Tempo is indicated to be unbounded [1] hence justifying no further processing or normalisation. To facilitate a more direct mapping between user input and song mood, the Google Emotions dataset, originally comprising 27 emotional features was streamlined into a core set of 8 distinct categories: Exuberant, Happy, Energetic, Frantic, Sad, Depression, Calm, and Contentment. This simplification involved strategically merging certain features based on their semantic similarity [14]. This transformation aimed to establish a clear correspondence between the predicted emotion of user text and the corresponding mood label associated with each song.

3.2 Music Classification

This section of the paper deals with the classification of the songs in our dataset. The classification method was adapted from Nuzzolo [1]. This paper investigated the link between audio features and perceived moods. Acoustical analysis was performed on 44,100 songs. From this insight Nuzzolo were able to determine that Timbre, Energy, Key and Tempo can be used to class songs into one of eight moods (Happy, Exuberant, Energetic, Frantic, Sad, Depression, Calm, Contentment).

Table 1: Moods Classified According to Musical Components (Derived from Nuzzolo [1])

Mood	**Energy**	**Timbre**	**Key**	**Tempo**
Happy	Medium	Medium	Very High	Very High
Exuberant	High	Medium	High	High
Energetic	Very High	Medium	Medium	High
Frantic	High	Very High	Low	Very High
Sad	Medium	Very Low	Very Low	Low
Depression	Low	Low	Low	Low
Calm	Very Low	Very Low	Medium	Very Low
Contentment	Low	Low	High	Low

Table 2: Mean values for audio features (Derived from Nuzzolo [1])

Mood	**Energy**	**Timbre**	**Key**	**Tempo**
Happy	0.2055	0.4418	967.47	209.01
Exuberant	0.317	0.4265	611.94	177.7
Energetic	0.4564	0.319	381.65	163.14
Frantic	0.2827	0.6376	239.78	189.03
Sad	0.2245	0.1572	95.654	137.23
Depression	0.1177	0.288	212.65	122.65
Calm	0.0658	0.1049	383.49	72.23
Contentment	0.1482	0.2114	756.65	101.73

3.2.1 Class Label Assignment

Nuzzolo's research paper provided insight to the mood of a song based on the values of the four attributes. Table 1 provides the categorical ranges for each feature per song mood. Table 2 provides the mean values for each category in Table 1. Using these two tables, fixed ranges were estimated for each category for each mood cluster. These ranges were further tested against a list of songs with pre-determined mood ratings. The boundary conditions were calibrated and finalised and given by Table 3.

With these finalised ranges the ideal cases from our data set were extracted. This was done by directly comparing each feature value of the song to predetermined ranges per feature. Songs whose feature values perfectly match within the ranges for a specific mood were given an appropriate label (Data imputation). However, due to the limited size of the dataset there were not enough ideal cases to properly train a model. With there being limited points, the ideal cases were also unbalanced having a small number of depression songs. To solve this problem, we implemented a method of generating new data points.

Table 3: Calibrated Ranges for each feature

Mood	**Energy**	**Timbre**	**Key**	**Tempo**
Happy	0.4 - 0.6	0.3 - 0.5	0.7 - 1.0	175+
Exuberant	0.6 - 0.8	0.3 - 0.5	0.6 - 0.7	160 - 175
Energetic	0.8 - 1.0	0.3 - 0.5	0.4 - 0.6	160 - 175
Frantic	0.6 - 0.8	0.6 - 1.0	0.2 - 0.4	175+
Sad	0.4 - 0.6	0.0 - 0.2	0.0 - 0.2	75 - 100
Depression	0.2 - 0.4	0.2 - 0.3	0.2 - 0.4	75 - 100
Calm	0.0 - 0.2	0.0 - 0.2	0.4 - 0.6	0 - 75
Contentment	0.2 - 0.4	0.2 - 0.3	0.6 - 0.7	75 - 100

3.2.2 Data Generation

By using the final ranges from 3.3.1 and random selection, new ideal samples for each of the eight moods were generated. These synthetic data points were combined with the ideal cases that were extracted. Points were generated so that each mood had 70 data points to represent them.

3.2.3 Support Vector Classification

Figure 1 depicts the clustering of songs based on the ranges of the acoustic features of songs established in section 3.3.1. The diagram also visually reveals the gaps between these ranges. Informed by these visualisations, a Support Vector Machine (SVM) was identified as the most suitable model for song classification. SVM is a supervised machine learning algorithm adept at both classification and regression tasks. Its core functionality lies in identifying the optimal decision boundary between potential outputs. In the context of our project, SVM excels at maximising the separation margins between data points, making it an ideal choice. Leveraging the ideal cases alongside the generated points, the SVM model was trained using feature values as the X vectors and mood labels as the Y vectors.

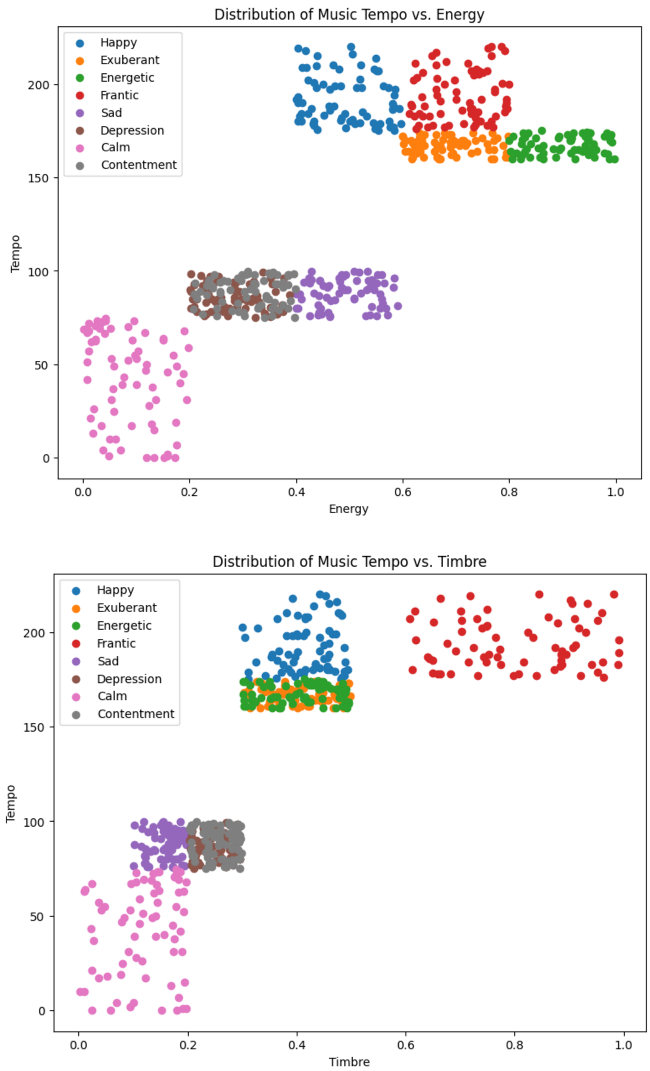

Figure 1: Diagram showing clustering of songs based on ranges assigned per mood category

3.3 Emotion Detection

3.3.1 Logistic Regression Model + TF-IDF

Term Frequency Inverse Document Frequency (TF-IDF) vectorized the textual data in 3.2 to measure how important a term is within a document relative to a collection of documents.

$$TF(w\ ,\ d)\ =\ \frac{number\ of\ occurrences\ of\ a\ word\ in\ the\ document}{number\ of\ words\ in\ document\ d}$$

$$IDF(w)\ =\ log(\frac{Number\ of\ documents}{Number\ of\ documents\ with\ w})$$

$$TFIDF(w,d)\ =\ TF(w,d)\ *\ IDF(w)$$

This was then used to train a multinomial logistic regression model to detect the underlying emotion of user input.

$$p_j(x) = \frac{\sum_{j=1}^{J-1} \exp(\beta_{0j} + \beta_{1j}X_1 + \cdots + \beta_{pj}X_p)}{1 + \sum_{j=1}^{J-1} \exp(\beta_{0j} + \beta_{1j}X_1 + \cdots + \beta_{pj}X_p)} \quad \text{for } j = 1 \text{ to } J-1$$

3.3.2 Logistic Regression Model + TF-IDF + Ngram

To further enhance the Multinomial Logistic Regression model described above, we incorporated N-grams into the feature set. N-grams represent sequences of n consecutive words. For instance, the sentence *"my cat ran away"* constitutes a 4-gram. During model training, we experimented with a combination of N-gram ranges, specifically n {1,2,3,4}. This exploration aimed to optimise the precision score achieved by the Logistic Regression model on the emotions dataset.

3.3.3 Recurrent Neural Network (RNN) using Long Short-Term Memory (LSTM)

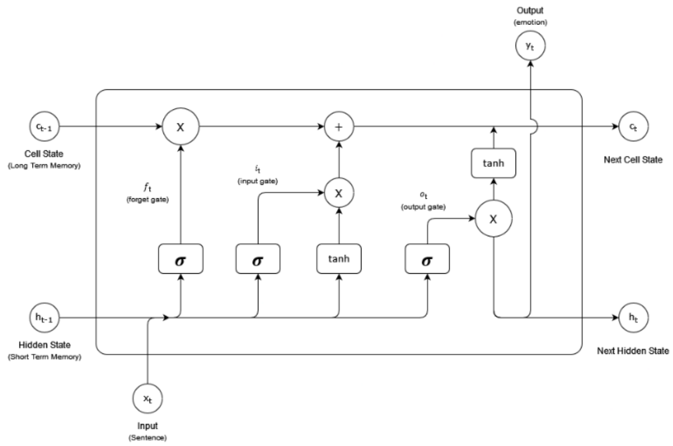

Figure 2: Diagram of an LSTM Cell used in the RNN for emotion detection from text

The Long Short-Term Memory (LSTM) network operates through a series of sequential steps to detect emotions from text. Initially, the input text is encoded into a numerical format suitable for processing by the LSTM. This encoding typically involves techniques like word embeddings, where each word is represented as a high-dimensional vector.

As the encoded text is fed into the LSTM, it proceeds sequentially, processing one word at a time in the order they appear in the text. At each time step, the LSTM updates its internal memory cell state (ct) and produces an output hidden state (ht). The memory cell state serves as a repository of long-term information, while the hidden state captures pertinent information for the current prediction.

Key to the LSTM architecture are its gate mechanisms: the forget gate, input gate, and output gate. These gates regulate the flow of information within the network. The forget gate (ft) decides which information to discard from the cell state, the input gate (it) determines how the incoming vector (xt) alters the cell state, and the output gate (ot) gives the detected emotion based on the current cell state. These gates are governed by sigmoid and tanh activation functions, allowing them to regulate the flow of information effectively.

As the LSTM processes the text, it updates its memory cell state and hidden state based on the input and the gate

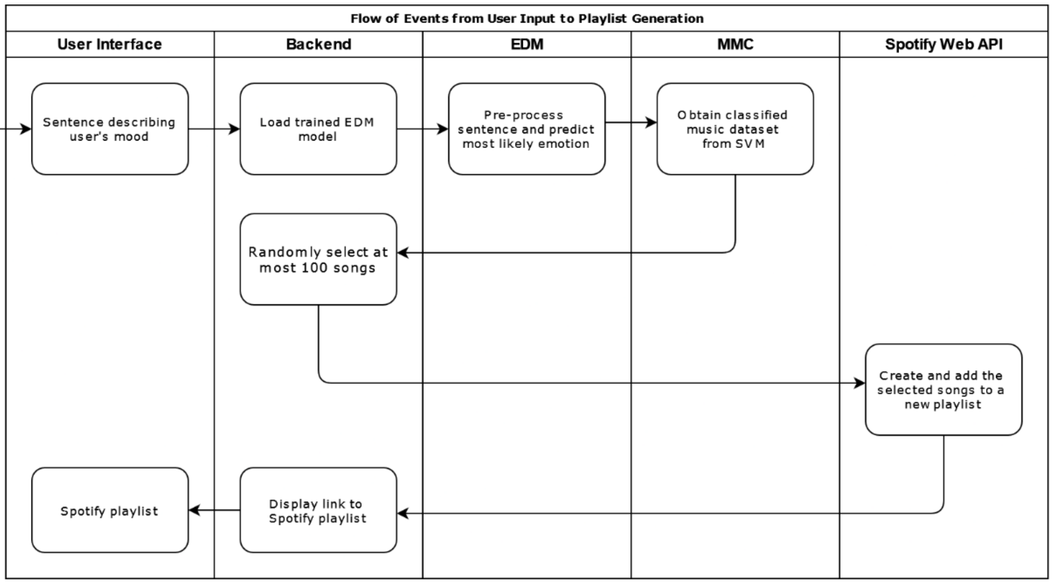

Figure 3: Diagram of Flow of Events from User Input to Spotify Playlist Generation

mechanisms. Once the entire sequence is processed, the final hidden state (or sequence of hidden states) is passed through a fully connected layer followed by a SoftMax activation function. This produces a probability distribution over emotion classes, with each class representing a specific emotion category.

3.4 Playlist Generation

The following diagram outlines the general process, from start to completion, of the system's playlist generation sequence. The user text is categorised to a mood label by the LSTM. Once the mood is identified from the text, a list of songs is dynamically generated by matching the detected mood with the mood labels attached to the list of classified songs. One hundred songs are randomly selected from the list and a Spotify playlist is generated through the Spotify Web API. The link to that playlist is then returned to the user.

4. RESULTS AND DISCUSSION

This section delves into the outcomes and insights garnered from the exploration of the predictive abilities of a combined SVM and LSTM model for mood-centric playlist generation. The study set out to tackle the challenge of precisely identifying emotional content from textual inputs, a pivotal aspect in the creation of mood-centric playlists derived from input text. Moreover, the objective was to effectively classify songs into the eight distinct moods, ensuring the selection of relevant songs aligned with the designated mood categories.

4.1 Evaluation Metrics and Scores

This section delves into the assessment metrics applied to both the SVM and LSTM. The SVM primarily relied on the Accuracy metric due to the balanced nature of the training dataset. The accuracy metric is given:

$$\frac{TP + TN}{(TP + TN) + (FP + FN)}$$

Furthermore, precision was used to evaluate the LSTM because the dataset used to train the model was unbalanced. The precision metric is given:

$$\frac{TP}{TP + FP}$$

Table 4: Metric scores for each model used

Model	Metric	Metric Score
SVM	Accuracy	0.91
LSTM	Precision	0.67

4.2 Comparative Analysis of EDMs

The following table gives the Accuracy and Precision metrics for the three best performing models used in the emotion detection section of this paper.

Table 5: Precision and Accuracy Metrics for the 3 Models

Model/ *Metric*	**Logistic Regression + TF-IDF**	**Logistic Regression + TF-IDF + Ngram (2,3)**	**LSTM**
Precision	0.611	0.522	0.671
Accuracy	0.612	0.390	0.537

The Logistic Regression model augmented with TF-IDF demonstrates at 0.611 precision, indicating its ability to accurately identify instances of specific emotions out of all classifications made by the model. Furthermore, this model achieved an accuracy score of 0.612, denoting its capability to correctly classify instances across all emotions.

Conversely, the introduction of N-grams (2,3) to the Logistic Regression + TF-IDF model results in a noticeable decline in performance. While the precision remains relatively high at 0.522, there was a significant decrease in accuracy, plummeting to 0.390. This suggests that while the inclusion of N-grams may capture certain linguistic nuances, it also introduces complexity that adversely affects the model's overall accuracy.

Finally, the LSTM model demonstrates competitive performance, with a precision score of 0.671 and an accuracy score of 0.537. In addition to surpassing the precision values of the Logistic Regression models, the LSTM model exhibits comparable accuracy, indicating its efficacy in correctly classifying instances across different emotions.

4.3 Peer Evaluation

Apart from standard evaluation metrics such as accuracy and precision, peer review was implemented to gauge the accuracy of the playlist from a human's perspective. Reviewers were asked to rate playlists that were generated for each of the eight moods. Reviewers were also given a playlist generated from their own input (indicative of the individual's mood) as well as a playlist of randomly selected songs. The reviewer would not know which playlist was generated by our model and which was randomly selected. After listening to the first 10 songs reviewers were asked to compare both playlists to the intended mood of the input on a Likert scale from 1 to 5.

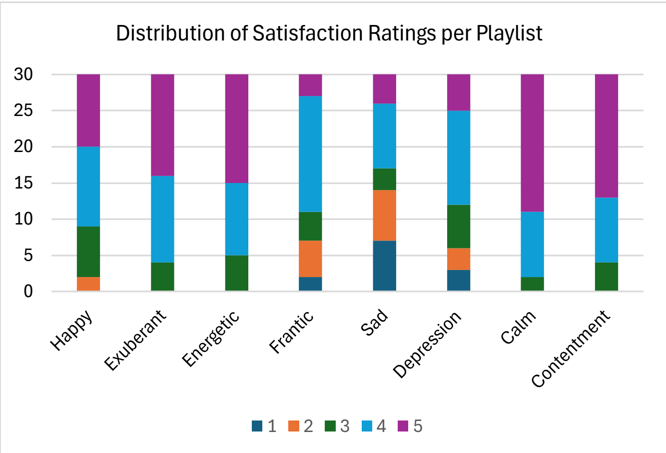

Figure 4: Satisfaction ratings generated from peer evaluation

Using the MMC, playlists were generated for the eight moods: Happy, Exuberant, Energetic, Frantic, Sad, Depression, Calm, Contentment. Participants were given a form to evaluate each playlist's effectiveness in representing the intended mood on a scale from 1 to 5. Fig 4 shows the results for this. Participants scored the playlist for Calm, Contentment and Exuberant very highly with a satisfaction rating of 92%, 88% and 87% respectively. However, the playlist generated for the mood Sad scored a satisfaction rating of 42%.

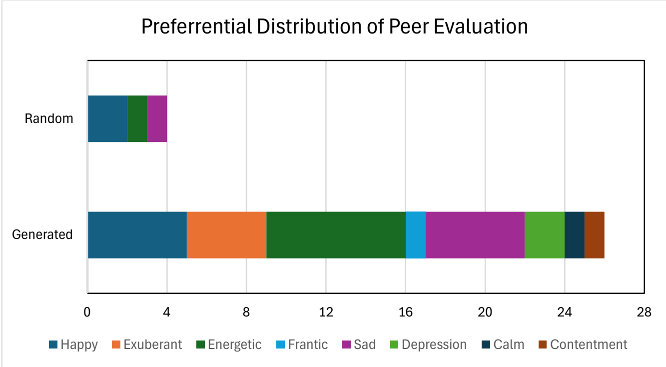

Figure 5: Approval ratings generated from peer evaluation

Additionally, when participants were asked to choose between a random playlist and the generated playlist, an 86% approval rating was seen in favour of our model (Seen in Fig 5).

4.4 Top Frequent Terms per Emotion Category

During emotion detection in text, analysing the Top Frequent Terms using TF-IDF for each emotion provides valuable insights into the prevalent themes and language patterns associated with each emotional state. This approach allows for the identification of key terms that strongly correlate with specific emotions, shedding light on the linguistic nuances underlying different emotional expressions. By examining the most frequent terms within each emotion category, patterns emerge, revealing the vocabulary most indicative of joy, sadness, excitement, calmness, and other emotional states. These findings not only enhance our understanding of the textual cues indicative of different emotions but also inform the development of more accurate and nuanced emotion detection algorithms tailored to diverse linguistic contexts.

Table 6: Top 15 Frequent Terms in the Google Emotions dataset collated using TF-IDF

Emotion	**Top 15 Frequent Terms**
Calm	thank, relaxed, good, satisfying, lucky, safe, god, relieved, finally, worry, better, relief, cool, relax, glad
Contentment	best, appreciated, thankful, thankfully, good, bless, congratulation, glad, congrats, luck, welcome, appreciate, great, thank, thanks
Depression	missed, wrong, grief, mistake, bad, death, shame, apologise, rip, guilty, dead, apology, died, regret, sorry
Energetic	yes, cake, favourite, new, interested, cheer, exciting, happy, amazing, birthday, awesome, wait, wow, interesting, excited
Exuberant	ridiculous, hahahaha, joking, kidding, laughing, laughed, hahaha, joke, laugh, fun, hilarious, funny, lmao, haha, lol
Frantic	cringe, nightmare, terrified, horrifying, dangerous, scare, creepy, worried, terrifying, fear, scary, horrible, afraid, terrible, scared
Happy	smile, sweet, cheer, awesome, favourite, haha, love, joke, enjoying, funny, fun, enjoyed, glad, enjoy, happy

Sad	heartbreaking, broken, lost, saddest, pain, unfortunately, sick, poor, miss, hurt, painful, bad, sadly, sorry, sad

5. LIMITATIONS

5.1 Limited Dataset Size and Timbre Values

Although a large dataset of 1.2 million data points was retrieved, it lacked a key attribute (timbre value) that restricted the dataset's completeness and variety. The constraint of only being able to scrape a limited number of data points per day to remedy this setback hindered the scalability and comprehensiveness of the Spotify dataset, which may have affected the model's performance and generalisability.

5.2 Issues with Emotions Dataset

The dataset sourced from Kaggle containing emotions data proved to be unsuitable for research due to its vulgar and unclear nature. This limitation likely affected the quality and reliability of the emotional labels assigned to the text data, compromising the accuracy and interpretability of the EMD.

5.3 Missing Lyrics for Spotify Tracks

Due to the absence of lyrics for the Spotify tracks in the Spotify dataset, sentiment analysis on the lyrical content could not be performed. Integrating the results of sentiment analysis on the song's lyrics may have enhanced the classification ability of the MMC by providing additional textual features related to the emotional content of the songs [11].

5.4 Subjectivity of Moods and Emotions

One of the inherent challenges in analysing and classifying moods and emotions analytically is their subjective nature. Different individuals may interpret, and express emotions differently based on various factors such as cultural background, personal experiences, and psychological disposition. This subjectivity introduces ambiguity and variability in the labelling and classification of emotional states, making it challenging to establish a standardised and universally applicable classification framework. The lack of objective criteria for defining and categorising emotions could lead to inconsistencies and discrepancies in the classification results obtained from the SVM and LSTM models, limiting its reliability and robustness in real-world applications.

6. FUTURE WORK

Moving forward, several avenues for improvement and expansion emerge based on the insights gained and limitations encountered during this research endeavour.

6.1 Data Augmentation and Diversification

Firstly, addressing the scarcity of timbre values and the constraints imposed by data collection limitations presents an opportunity for data augmentation and diversification. Augmentation techniques, such as synthetic data generation or integration from alternative sources, could enrich the dataset's variety and quantity, enhancing the robustness and generalizability of the model.

Secondly, to overcome the inadequacies of the emotions dataset obtained from Kaggle, meticulous preprocessing and filtering procedures are warranted. By curating or annotating a new emotions dataset aligned with the research objectives, the quality and appropriateness of emotional labels assigned to textual data can be ensured, thereby fostering more accurate and reliable emotion detection outcomes.

6.2 Sentiment Analysis on Song Lyrics during Music Mood Classification

Integrating lyrics data for Spotify tracks and conducting sentiment analysis on the lyrical content offer avenues for enhancing the model's classification performance. Leveraging existing lyrics databases or APIs to retrieve lyrics, coupled with sentiment analysis algorithms, can extract emotional features from the lyrical text, enriching the input features and improving emotion detection accuracy.

6.3 Use Expert-annotated Emotion Datasets

Additionally, the incorporation of human annotation or expert feedback is crucial for validating emotional labels and refining the dataset's quality. By soliciting annotations from human raters or consulting domain experts in psychology or emotional analysis, the accuracy and consistency of emotional labels can be verified, enhancing the model's reliability and interpretability.

6.4 Supplementary Models

Using Latent Dirichlet Allocation (LDA) during training of the EDM may add greater insight to the model and thus further enhance the performance of the model.

7. CONCLUSION

This research presented a system capable of generating mood-specific music playlists based on user input. The system leverages a Support Vector Machine (SVM) for music mood classification and a Long Short-Term Memory (LSTM) network for emotion detection from textual input. The LSTM network exhibited promising performance in emotion detection, achieving a precision of 0.67. User evaluations further confirmed the effectiveness of the generated playlists in capturing the intended mood.

While limitations such as dataset size and the inherent subjectivity of emotions were acknowledged, the research paves the way for future advancements. Data augmentation techniques and sentiment analysis of song lyrics hold promise for enhancing the model's accuracy and robustness. Additionally, incorporating expert-annotated emotion datasets and exploring supplementary models like Latent Dirichlet Allocation (LDA) present avenues for further exploration.

Overall, this research contributes to the fields of music mood classification and emotion detection from text. The proposed application has the potential to revolutionise music recommendation systems and personalise playlist generation, offering users a tailored musical experience that resonates with their emotional state.

8. REFERENCES

[1] M. Nuzzolo, "Music Mood Classification | Electrical and Computer Engineering Design Handbook," *Tufts.edu*, 2013. https://sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/ (accessed Mar. 15, 2024).

[2] Spotify Technology S.A., "Web API | Spotify for Developers," *developer.spotify.com*, 2024. https://developer.spotify.com/documentation/web-api (accessed Mar. 16, 2024).

[3] Y. Liang and M. Willemsen, "Interactive Music Genre Exploration with Visualization and Mood Control," Apr. 2021. Accessed: Mar. 16, 2024. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/3397481.3450700

[4] G. Mitri, A. L. Uitdenbogerd, and V. Ciesielski, "Automatic Music Classification Problems," in *Conferences in Research and Practice in Information Technology*, Australian Computer Society Inc, 2004, pp. 315–322. doi: https://doi.org/10.5555/979922.979960.

[5] S. Bhattacharya, A. Agarwala, and S. Roy, "Mood Detection and Prediction Using Conventional Machine Learning Techniques on COVID19 Data," *Social Network Analysis and Mining*, vol. 12, no. 1, Sep. 2022, doi: https://doi.org/10.1007/s13278-022-00957-x.

[6] M. R. Dalida, L. Bianca Aquino, W. C. Hod, R. Ann Agapor, S. L. Huyo-a, and G. Avelino Sampedro, "Music Mood Prediction Based on Spotify's Audio Features Using Logistic Regression," *2022 IEEE 14th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Dec. 2022, doi: https://doi.org/10.1109/hnicem57413.2022.10109396.

[7] T. Dang and K. Shirai, "Machine Learning Approaches for Mood Classification of Songs toward Music Search Engine," presented at the 2009 International Conference on Knowledge and Systems Engineering, Oct. 2009. doi: https://doi.org/10.1109/kse.2009.10.

[8] S. P. Robbins and T. A. Judge, *Organizational Behavior, Global Edition*. Pearson Higher Ed, 2023.

[9] C. Laurier, J. Grivolla, and P. Herrera, "Multimodal Music Mood Classification Using Audio and Lyrics," *IEEE Xplore*, Dec. 01, 2008. https://ieeexplore.ieee.org/abstract/document/4725050 (accessed Mar. 17, 2024).

[10] G. Agarwal, S. Gupta, S. Agarwal, and A. K. Rai, "Emotion classification for musical data using deep learning techniques," *International Journal of Reconfigurable and Embedded Systems (IJRES)*, vol. 12, no. 2, p. 240, Jul. 2023, doi: https://doi.org/10.11591/ijres.v12.i2.pp240-247.

[11] M. Van Zaanen and P. Kanters, "Automatic Mood Classification Using TF*IDF Based on Lyrics. AUTOMATIC MOOD CLASSIFICATION USING TF*IDF BASED ON LYRICS," 2010. Accessed: Mar. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/220723672_Automatic_Mood_Classification_Using_TFIDF_Based_on_Lyrics

[12] S. Hizlisoy, S. Yildirim, and Z. Tufekci, "Music emotion recognition using convolutional long short term memory deep neural networks," *Engineering Science and Technology, an International Journal*, vol. 24, no. 3, pp. 760–767, Nov. 2020, doi: https://doi.org/10.1016/j.jestch.2020.10.009.

[13] M. Tan, D. S. Cicero, B. Xiang, and B. Zhou, "LSTM-based Deep Learning Models for non- Factoid Answer Selection," 2016. Accessed: Mar. 18, 2024. [Online]. Available: https://arxiv.org/pdf/1511.04108.pdf

[14] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A Dataset of Fine-Grained Emotions," *ACLWeb*, Jul. 01, 2020. https://aclanthology.org/2020.acl-main.372/ (accessed Mar. 18, 2024).

[15] S. Raschka, "MusicMood: Predicting the mood of music from song lyrics using machine learning," *ResearchGate*, Nov. 2016. https://www.researchgate.net/publication/309606883_MusicMood_Predicting_the_mood_of_music_from_song_lyrics_using_machine_learning