A blue and yellow crest with a gold crest and blue text

Description automatically generated

**CAPSTONE PROJECT 2**

**Final Report**

**VibeSync: An Emotion-Based Recommendation System with Integrated Chatbot**

by

LESTER KOON ZHY MIN

20068813

Bachelor of Science (Honours) in Computer Science

Supervisor: Muthukumaran Maruthappa

Semester: September 2024

Date: 7 January, 2025

Department of Smart Computing and Cyber Resilience (DSCCR)

School of Engineering and Technology

Sunway University

Table of Contents

[Abstract 5](#_Toc187156379)

[1.0 INTRODUCTION 5](#_Toc187156380)

[1.1 Introduction 5](#_Toc187156381)

[1.2 Problem Statement 7](#_Toc187156382)

[1.3 Aim & Objectives 7](#_Toc187156383)

[Aim 7](#_Toc187156384)

[Objectives 8](#_Toc187156385)

[2.0 LITERATURE REVIEW 8](#_Toc187156386)

[2.1 Review of Key Topics 8](#_Toc187156387)

[2.1.1 Effects of Music on Mood 8](#_Toc187156388)

[2.1.2 Chatbots in Emotion Detection 10](#_Toc187156389)

[2.1.3 Music Emotion Recognition 13](#_Toc187156390)

[**2.1.4 Machine Learning and AI in Music Recommendation** 18](#_Toc187156391)

[**2.2 Related Works** 20](#_Toc187156392)

[**2.2.1 "Emotion Based Music Playlist Recommendation System using Interactive Chatbot"** 21](#_Toc187156393)

[**2.2.2 "Emotune: Emotion And Gender Aware Music Generation Chatbot"** 23](#_Toc187156394)

[**2.2.3 "Emousic: Emotion and Activity-Based Music Player Using Machine Learning"** 25](#_Toc187156395)

[**2.2.4 "Induced Emotion-Based Music Recommendation through Reinforcement Learning"** 27](#_Toc187156396)

[**2.2.5 “Smart music player integrating facial emotion recognition and music mood recommendation”** 29](#_Toc187156397)

[**2.3 Findings and Results** 31](#_Toc187156398)

[**2.3.1 Comparative Analysis of Existing Systems** 34](#_Toc187156399)

[**2.4 Proposed Solution** 37](#_Toc187156400)

[3.0 Methodology 41](#_Toc187156401)

[3.1 Tools 41](#_Toc187156402)

[3.2 Data Collection and Processing 42](#_Toc187156403)

[3.3.1 Music and Emotion Data 42](#_Toc187156404)

[3.2.2 Chatbot Text Data 43](#_Toc187156405)

[3.3 Detect Emotion of Songs with Weighted Labels 45](#_Toc187156406)

[3.3.1 Valence Arousal Prediction 45](#_Toc187156407)

[3.3.2 Weighted Emotion Classification 50](#_Toc187156408)

[3.4 Chatbot Emotion Prediction 54](#_Toc187156409)

[3.5 Chatbot Implementation 56](#_Toc187156410)

[3.6 Website Development 59](#_Toc187156411)

[4.0 Results & Discussion 62](#_Toc187156412)

[4.1 Valence-Arousal Prediction 62](#_Toc187156413)

[4.1.1 Valence Prediction Results 62](#_Toc187156414)

[4.1.2 Arousal Prediction Results 64](#_Toc187156415)

[4.2 Weighted Emotion Classification 67](#_Toc187156416)

[4.3 Chatbot Emotion Prediction 69](#_Toc187156417)

[5.0 Conclusion 72](#_Toc187156418)

[References 75](#_Toc187156419)

**Table of Figure**

[Figure 1: Russell’s emotion circumplex and core emotions for each quadrant [36]. 17](#_Toc187156420)

[Figure 2: Thayer’s mood model and emotion categories [37]. 18](#_Toc187156421)

[Figure 3: Snippet of the processed MER database. 43](#_Toc187156422)

[Figure 4: Dataset emotion distribution before processing. 44](#_Toc187156423)

[Figure 5: Dataset emotion distribution after processing. 44](#_Toc187156424)

[Figure 6: Arousal prediction model comparison results. 48](#_Toc187156425)

[Figure 7: Emotion Quadrants established in Russell’s circumplex model [39]. 49](#_Toc187156426)

[Figure 8: Linear Regression model fit for valence scaling, showing the relationship between Spotify valence and scaled valence, with a 95% confidence interval. 63](#_Toc187156427)

[Figure 9: Residual diagnostics for the Linear Regression model, including Residuals vs. Fitted, Normal Q-Q Plot, Residual Distribution, and Scale-Location Plot. 64](#_Toc187156428)

[Figure 10: Predicted vs. actual arousal values for the training dataset, demonstrating the close alignment of predictions with the actual values. 65](#_Toc187156429)

[Figure 11: Residual diagnostics for the XGBoost model, including Residuals vs. Fitted and Residual Distribution. 66](#_Toc187156430)

[Figure 12: Relative importance of Spotify features in the arousal prediction task, highlighting the dominant role of Sp\_Energy. 66](#_Toc187156431)

[Figure 13: Predicted vs. actual arousal values for the test dataset, showcasing the model's generalization capabilities. 67](#_Toc187156432)

[Figure 14: Confusion matrix illustrating true vs. predicted emotion categories, highlighting areas of high accuracy and emotional overlap. 71](#_Toc187156433)

**Table of Tables**

[Table 1: Common Music Genres and Their Associated Moods 10](#_Toc187156434)

[Table 2: Techniques and Methods for Emotion Detection 12](#_Toc187156435)

[Table 3: Techniques for Recognizing Emotion in Music 15](#_Toc187156436)

[Table 4: Comparison of Key Features and Functionalities of Current Systems 34](#_Toc187156437)

[Table 5: Valence prediction model comparisons. 47](#_Toc187156438)

[Table 6: Arousal prediction model comparisons. 49](#_Toc187156439)

[Table 7: Emotions classified according to musical components [37] [40]. 53](#_Toc187156440)

[Table 8: Classification metrics for each emotion category, including precision, recall, F1-score, and support. 71](#_Toc187156441)

# Abstract

This study presents VibeSync, an innovative emotion-based music recommendation system integrated with a context-aware chatbot. The project leverages advanced AI, machine learning, and emotion recognition models to enhance user experience by aligning music recommendations with emotional states. By employing Valence-Arousal prediction, Weighted Emotion Classification, and a fine-tuned transformer-based emotion detection model for the chatbot, the system achieves nuanced emotion recognition and personalized music recommendations.

The methodology integrates multiple data sources, including Spotify features, to predict valence and arousal scores, classifying songs into emotional quadrants and assigning weighted emotion labels based on audio features. Additionally, the chatbot employs a fine-tuned DistilRoBERTa model to detect user emotions from text inputs, ensuring tailored conversational interactions. Results demonstrate the system's effectiveness, with accurate emotion classification for songs and high precision, recall, and F1 scores in chatbot emotion detection.

Despite challenges such as reliance on external APIs and computational demands, VibeSync represents a significant advancement in emotion-aware music recommendation systems. By bridging gaps in current systems, this project provides a more intuitive, emotionally resonant user experience, emphasizing the potential of AI-driven personalization in music consumption.

***Keywords: Emotion Detection, Music Recommendation, Valence-Arousal Model, Chatbot, Weighted Emotion Classification, Machine Learning***

# INTRODUCTION

## 1.1 Introduction

Music and song has long been established as a vital tool for community bonding and promoting camaraderie, as it has the ability to improve coordination and reducing tensions within a group [1]. Music serves as a form of communication that goes beyond cultural and language differences. Music indeed serves as a profound medium for emotional expression, influencing a wide range of feelings such as happiness, sadness, and relaxation. It can evoke a wide range of emotions, from joy to sorrow, tranquillity to excitement in listeners, which varies based on individual experiences and contexts [2]. The psychological effects of music are such that it can enhance mood, reduce stress, and even alleviate symptoms of depression. Research has demonstrated that music stimulates diverse areas of the cerebral cortex, encompassing those engaged in affect, recollection, and corporeal movement, which is why it is often used to aid in cognitive therapy and rehabilitation processes [3]. Nair [4] supports this relationship in their study, stating that it is strongly evident that music contains therapeutic properties by stimulating the body and mind, and is a proven remedy to mental ailments and boosting self-esteem. This association constitutes the cornerstone for modern music recommendation algorithms and systems, which aims to personalize the user experience by aligning song selections with the user’s preferences. Nevertheless, prevailing systems predominantly depend on metadata or user listening history, often neglecting the intricate emotional connections music fosters​. Additionally, streaming applications and platforms, such as Spotify and Apple Music, which have a dominant monopoly on the music application market, have completely revolutionized music consumption, which has led to popularity overshadowing quality and accurate recommendations. This in turn significantly the ability of users to find more relatable music and discover newer artist [5].

Chatbots are widely used in business applications such as flight booking and FAQ agents. They can also be utilized to helping individuals manage their emotions without external assistance, through emotion recognition and solution recommendation systems, allowing the bots to simulate one-on-one interactions and recommend ways to handle emotions effectively. In recent years, the advent of chatbots and artificial intelligence (AI) has revolutionized the way we interact with music, leading to the development of sophisticated music recommendation systems. However, despite the advancements in machine learning algorithms and music recommendation technologies, existing systems frequently fall short of entirely encapsulating the intricacy and subtlety of human emotions [6] [7]. Therefore, there is a significant need for an emotion-based music recommendation system that can accurately recognize and respond to user emotions, offering a more personalized and emotionally resonant music experience. Factors such as the level of personalization, the complexity of the conversational interactions, and the integration with other systems are all vital in determining the viability and effectiveness of the chatbot [8].

VibeSync aims to address these limitations by developing an emotion-based music recommendation system integrated with a chatbot for seamless interaction. The project leverages state-of-the-art AI and machine learning techniques to classify songs based on weighted emotion labels, enabling a more nuanced understanding of user needs. This dynamic approach ensures that music suggestions are not only based on the user's preferences but are also tailored to their current emotional needs, fostering a more meaningful connection between the listener and the music. VibeSync will be able to overcome challenges of its predecessors in music recommendation systems, which struggled to consider the emotional aspect of music and struggle to adapt to greatly varying individual experiences and contexts in real-time.

## 1.2 Problem Statement

Despite the significant advancements in music streaming and recommendation technologies, current systems often fail to capture and respond to music's emotional nuances and users' emotional states. Traditional recommendation methods, such as collaborative and content-based filtering, primarily focus on user preferences and song characteristics without adequately considering the emotional context. This causes disconnect between the user’s emotional needs and the recommended music. Additionally, dynamically adjust recommendations based on real-time emotional feedback from users, add an extra layer of complexity that current system have fail to implement. Furthermore, the dominant popularity of mainstream music can restrict the discovery of emotionally resonant or lesser-known artists, restricting the variety and significance of suggestions. Consequently, there is a clear and pressing need for an emotion-aware music recommendation system capable of understanding and responding to the user's emotional state, providing music that is not only tailored to their preferences but also to their emotional needs at any given time.

## 1.3 Aim & Objectives

### Aim

The aim of this project is to develop VibeSync. VibeSync is an advanced emotion-based music recommendation system with an integrated chatbot that enhances user experience by providing personalized music recommendations which aligns with the user's emotional state. The system will classify songs with weighted emotion labels and use these classifications to generate dynamic playlists that adapt to the user’s mood, which is respectively detected through an AI-powered chatbot. VibeSync will be able to continuously monitor the user’s detected mood and refine its recommendations, accordingly, creating a more intuitive and personalized listening experience.

### Objectives

The main objectives of VibeSync are:

* Detect the Emotions of Songs with Weighted Emotion Labels

Develop and implement a system that detects and classifies the emotional content of songs by assigning weighted emotion labels based on audio features.

* Create a Recommendation System

Build a recommendation engine that suggests music based on the user’s emotional state and listening preferences.

* Create a Context-Aware Chatbot that can Detect User's Emotions with Weights

Develop and train a chatbot to understand context and detect the user’s emotional state while maintain a conversation.

* Dynamically Recommend Songs Based on Detected Mood

Implement a dynamic recommendation system that continuously adapts to the user’s changing emotional state, using real-time mood detection.

# 2.0 LITERATURE REVIEW

## 2.1 Review of Key Topics

### 2.1.1 Effects of Music on Mood

Music can affect our mood, regardless of our ability to identify or reproduce the specific melodies and rhythm. Music profoundly influences the emotions of humans and can evoke a wide range of emotional states, from happiness and excitement to sadness and sorrow. Hence, it becomes imperative to understand how music affects moods, for creating tailored music suggestions that aim to enhance user experiences by aligning music choices with emotional states.

#### Influence of Music on Emotional States

Research in psychology and neuroscience has consistently demonstrated that music is a tool that impacts the human mood and behaviour, regardless of what kind of music it is, it affects in either good or bad ways. For instance, according to previous studies, music's impact on mood is largely mediated by its structural elements such as tempo, rhythm, and melody [2]. Fast tempos and major keys are generally associated with positive emotions, whereas slow tempos coupled with minor keys tend to evoke negative emotions. The ability of music to influence mood is leveraged in therapeutic settings, where music therapy is used to improve mental health and emotional well-being [4].

The psychological mechanisms underlying music's impact on mood involve complex interactions between auditory perception, emotional processing, and memory. Neurologically, music activates multiple brain regions which are linked with the motivation and reward circuitry, such as the amygdala and hippocampus [3]. The activation can lead to changes in blood pressure and heart rate, which ascertains why certain pieces of music can trigger vivid emotional responses [9].

#### Empirical Studies on Music and Mood

Various empirical studies have investigated the specific effects of music on mood. Nawaz and Rana [2] conducted an empirical investigation that found that observed listening to music with different genres and tempos have distinct effects on mood states. It is revealed that participants who listened to slower, more melancholic music reported increased feelings of sadness and reflection, while those who listened to upbeat music reported increased levels of happiness and energy [10].

|  |  |
| --- | --- |
| Music Genre | Mood Induced |
| Pop/Rock | Happiness, Excitement |
| Classical | Relaxation, Calmness |
| Jazz | Creativity, Relaxation |
| Blues | Sadness, Reflection |
| Electronic | Energy, Motivation |
| Music Genre | Mood Induced |

Table 1: Common Music Genres and Their Associated Moods

Another study by Dalida et al. [11], provided evidence that certain auditory characteristics, such as valance, energy, tempo and danceability are strong predictors of the emotional content of music tracks.

In essence the relationship between music and emotions is intricate, involving complex interactions between psychological and empirical factors. Drawing insights from previous studies is vital for the development of emotion-based music recommendation systems. By incorporating mood prediction algorithms, these systems can curate playlists that match or influence the user’s current emotional state, thereby boosting the overall listening experience and supports emotional well-being.

### 2.1.2 Chatbots in Emotion Detection

The rise of artificial intelligence and automated technologies have made chatbots have become an integral part of modern user interaction systems in various industries. However, until recent years, chatbots and machines lack many important social and emotional skill to replicate natural human communication. Thus, emotion detection in chatbots involves various sophisticated techniques to analyze text, voice, or other inputs to determine the emotional tone of the user.

#### Role of Chatbots in Emotion Detection

Chatbots play a pivotal role in detecting emotions to provide more empathetic and personalized user interactions, by effectively communicating and adapting to create natural human interaction. They are crucial in applications such as customer service, mental health services, and personalized content generation. In the context of emotion detection, chatbots leverage the use of various natural language processing (NLP) and machine learning algorithms to understand and respond to human emotions, enhancing user interaction by understanding and responding to emotional states [12].

#### Techniques and Methods for Emotion Detection in Chatbots

Several methods have been developed for emotion detection using chatbots. Emotion detection in chatbots primarily relies on natural language processing (NLP) and machine learning models. For example, Karna et al. demonstrated that LSTM-based models significantly outperform traditional methods recognizing emotions from text inputs [13]. These models can learn complex patterns in text data, allowing chatbots to accurately identify a range of emotions, from happiness and excitement to sadness and anger.

The process typically involves several steps:

* Text Preprocessing: Cleaning and processing the text data by eliminating noise, normalizing, and tokenizing.
* Feature Extraction: Extracting features such as contextual information, emotional keywords, tone, and sentiment scores.
* Model Training: Training machine learning models on labeled datasets to recognize and predict different emotional states accurately.
* Emotion Classification: Classifying the user input into a set of predefined emotion categories.

Additionally, these techniques enable chatbots to analyze and interpret user inputs effectively.

|  |  |
| --- | --- |
| Technique | Description |
| Natural Language Processing (NLP) | Analyzes text input to detect emotions. Common methods include sentiment analysis and emotion classification. |
| Machine Learning Models | Utilizes models such as Long Short-Term Memory (LSTM) networks for handling sequential data and improving emotion detection accuracy [13]. |
| Deep Learning Approaches | Employs deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to capture complex patterns in text and voice inputs [13]. |
| Transfer Learning | Transfers the learned knowledge, such as features, models, or representations, from a source domain to a target domain [14]. |

Table 2: Techniques and Methods for Emotion Detection

1. Natural Language Processing (NLP)

* **Sentiment Analysis**: This technique classifies text as positive, negative, or neutral to gauge user sentiment. Sentiment analysis often uses supervised learning algorithms such as Support Vector Machines (SVM) and Naive Bayes classifiers trained on labeled datasets [13].
* **Emotion Classification**: More advanced than sentiment analysis, emotion classification identifies specific emotions in text such as joy, sadness, anger, or fear. This can be done using lexicon-based approaches, where words are matched with pre-defined emotion dictionaries, or using machine learning models that classify emotions based on text features [13].

1. Machine Learning Models

* **k-Nearest Neighbors (k-NN):** This algorithm classifies a data point based on how its neighbors are classified. It is a simple, instance-based learning method used for various classification tasks, including emotion detection. Although k-NN is straightforward and easy to implement, it may not always be the most efficient for large datasets due to its computational complexity during the classification phase [7].
* **Long Short-Term Memory (LSTM):** LSTMs are a type of recurrent neural network (RNN) capable of learning order dependence in sequence prediction problems. This makes them highly effective for tasks involving sequential data, such as text emotion recognition. LSTM models can capture the context and sequential dependencies in user input, providing accurate emotion detection [9].
* **Random Forests:** This ensemble learning method operates by constructing multiple decision trees during training and outputting the mode of the classes for classification tasks. Random Forests are robust to overfitting and can handle large datasets with higher accuracy, making them suitable for emotion detection tasks [7].

1. Deep Learning Approaches
   * **Convolutional Neural Networks (CNNs):** CNNs are particularly effective for processing grid-like data structures, such as images or sequences, making them suitable for identifying patterns in text or audio inputs related to emotions. CNNs are used to capture local patterns and hierarchies in the input data, improving the detection accuracy of nuanced emotional cues [15].
2. Deep Learning Approaches
   * **Inductive Transfer Learning:** Inductive learning is a type of transfer learning where the source and target tasks are different, but the domains can be the same. This is mainly used to improve the efficiency and effectiveness of the target task by increasing its knowledge from the source task, as it is already familiar the data structure [16].
   * **Transductive Transfer Learning:** Transductive learning is the inverse of inductive learning, as the source and target tasks are the same, but the domains are different. This method is mainly utilized in the scenario where the target domain has little or no labelled data. This allows the target domain model to better recognize patterns as the intent of the model is similar [17].
   * **Unsupervised Transfer Learning:** Unsupervised learning or zero-shot learning is when both the source and target task transfer knowledge without using any labelled data. This is especially useful to develop new knowledge by generalizing a large pool of data [18].

#### **Challenges and Limitations**

Despite the advancements in emotion detection, several challenges and limitations remain:

* **Ambiguity in User Input:** Interpreting user input can be challenging due to ambiguities and variations in common language. Contextual nuances, sarcasm and slang may be misinterpreted and lead to a loss of accuracy when detecting emotions [12].
* **Context Dependency:** The significance of emotions may be language-specific and chatbots need to get this right to throw responses back in a correct way. These demands elaborate models capable of understanding the context [12].
* **Quality of Training Data:** The performance of emotion detection models can be heavily influenced by the diversity and quality of training data. Data has biases that can inaccurately detect or predict emotions, leading to random responses [12].
* **Real-Time Processing:** Chatbots have to process and reply to user inputs in real-time, which requires efficient and fast models. Ensuring low latency while maintaining high accuracy is a technical challenge [19].

In the context of emotion-based music recommendation systems, chatbots are vital in reflecting the user’s current mood and emotions through interaction. This would enable the system to accurately recommend music that aligns with or aims to improve the user's mood. This connection improves the customization of the music recommendation system, making it more sensitive to the user's emotional requirements, achieving the project’s goal.

### 2.1.3 Music Emotion Recognition

Recognizing the emotion or mood in music is the main foundation in developing personalized music recommendation systems. The ability to accurately identify emotions conveyed by musical pieces allows the system to tailor recommendations to users' emotional states, thereby enhancing user satisfaction and engagement.

#### Techniques for Recognizing Emotion in Music

Various techniques are used to recognize emotions in music, each with its own advantages and challenges. The table below summarizes these techniques, highlighting their strengths and limitations.

|  |  |  |
| --- | --- | --- |
| Techniques | Advantages | Challenges |
| Audio-Based Emotion Recognition | Captures acoustic properties effectively | Requires sophisticated models for complex compositions and dependent on song segment used |
| Lyrics-Based Emotion Recognition | Analyzes textual content for emotional cues | Limited to songs with lyrics and may not understand lyrical context perfectly |
| Multimodal Emotion Recognition | Combines audio and lyrics for comprehensive analysis | Requires integration of diverse data types |

Table 3: Techniques for Recognizing Emotion in Music

#### Audio-Based Emotion Recognition

Audio-based emotion recognition relies on analyzing the acoustic properties of music. Techniques such as Mel-Frequency Cepstral Coefficients (MFCCs), chroma, spectral contrast, tempo, and rhythm analysis are commonly extracted to provide emotional context to audio segments [20]. These features capture the semantic content and emotional tone of the lyrics which are then processed, which may include feature scaling, segmentation, and normalization.

Machine learning models, such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and deep learning architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are trained on the preprocessed features. These models will learn the relationship between emotions and their corresponding audio features, then gaining the ability to predict the emotional content of new audio samples [21] [22].

#### Challenges and Limitation in Audio-Based Emotion Recognition

* Variability in Music**:** Different genres and styles of music can significantly impact the performance of emotion recognition models. For instance, classical music and heavy metal have different acoustic properties that can affect emotion detection accuracy [21].
* Subjectivity of Emotion: Emotions are inherently subjective and can vary from person to person. This variability makes it challenging to create models that can accurately predict emotions for all listeners [23].
* Complexity of Music Audio Signals: Music signals are complex and require sophisticated processing techniques to accurately capture the emotional nuances through acoustic features such as intensity, timbre, and rhythm [23].

#### Lyrics-Based Emotion Recognition

Words or lyrics-based emotion recognition involves recognizing the textual content of songs to determine their emotional tone. This can be done using natural language processing (NLP) techniques such as sentiment analysis, lexicon-based approaches, and deep learning models to extract meaningful patterns and sentiments from underlined in text or lyrics. The lyrics are cleaned and tokenized to remove stop words, punctuation, and other irrelevant elements. This step ensures that the text data is in a suitable format for further analysis [3]. Then, features such as term frequency-inverse document frequency (TF-IDF), word embeddings, and emotional lexicons are extracted from the lyrics [24]. Machine Learning model such as Random Forest, and K-Nearest Neighbors (KNN) will be trained to be able to assign a song to e distinct emotional categories, by using the NLP to extract and match keywords from all songs in its emotional category [3].

#### Challenges and Limitation in Lyrics-Based Emotion Recognition

* Ambiguity in Language: Lyrics often contain metaphors, slang, and ambiguous language, making it challenging to accurately interpret emotions which increases the complexity of NLP techniques needed. For instance, words such as “yeah” and “oh” can be interpreted in any emotional category and skew the accuracy of the final results [3].
* Contextual Understanding: Understanding the context in lyrics is crucial for accurate emotion detection but can be subjective and context dependent. This makes it difficult to capture complex emotional expressions based solely on text [25].
* Cultural Differences: Cultural variations can influence how emotions and storytelling elements expressed in lyrics can vary significantly across different languages and cultural contexts. This significantly effects the system's effectiveness and generalizability, if the model is trained entirely in English without consideration of other languages.

#### Multimodal Emotion Recognition

Multimodal emotion recognition fuses audio and lyrics analysis for optimal accuracy in detecting the encoded emotion in the music. This way, it combines the strengths from each modality to provide a pulse of the emotional content in the music. This statement is supported by evidence from the study of Delbouys et al., where they prove that multimodal deep models outcompete traditional models in this task of getting music emotion [20]. Different deep learning models are employed to enhance emotion recognition accuracy, especially when dealing with complex relationships between audio, lyrics, and emotions. Methods such as LSTM-based text emotion recognition and nested Long Short-Term Memory models are used to extract and classify emotions from different modalities to precisely and consistently recognize emotions [19]. However, challenges in complexity and synchronizations, especially in larger datasets severely affect the efficiency of this approach [20].

In conclusion, understanding the emotions conveyed in music plays a vital role in customizing music recommendations to align with the user’s emotional state. The systems can offer more personalized and emotionally impactful user interactions, by effectively recognizing the emotional elements in audio and lyrics of songs. Future research may focus on improving model accuracy and exploring new ways to combine audio and lyrical data effectively.

#### Music Emotion Classification Models

Psychology researchers have discussed for a long time how emotions can be represented and classified. This study has led to the proposal of several emotion taxonomies over the last century, which can be grouped into two major paradigms: categorical (or discrete) models and dimensional models [35].

Among these, Russell’s circumplex model of affect has attained recognition in the music emotion recognition field due to its simplicity and versatility. Russell's Valence-Arousal model is a widely used framework for describing emotions. It proposes that emotions can be represented along two orthogonal dimensions: valence, which refers to the pleasantness or unpleasantness of an emotion, and arousal, which refers to the intensity or activation level of an emotion [33]. Unlike categorical models, which classify emotions into discrete groups (e.g., happiness, sadness), Russell's model represents emotions along two continuous axes: valence and arousal.

* **Valence:** This dimension measures the positivity or negativity of an emotion. Positive valence is associated with pleasant emotions (e.g., joy, excitement), while negative valence corresponds to unpleasant emotions (e.g., sadness, anger).
* **Arousal:** This dimension captures the intensity or activation level of an emotion. High arousal indicates energetic emotions (e.g., anger, excitement), while low arousal represents calm or subdued emotions (e.g., contentment, sadness).

A diagram of different emotions

Description automatically generated

Figure 1: Russell’s emotion circumplex and core emotions for each quadrant [36].

Another popular model would be, Thayer's Model of Mood, which is a two-dimensional framework adapted from Russell’s circumplex. The model employs two primary dimensions: energy and stress, which align closely with the dimensions of arousal and valence in Russell's model​ [23]. However, Thayer’s model extends the basic two-dimensional framework to include eight categories, which encompass the extremes of the two dimensions (energy and stress) as well as their intersections. These categories allow for a more nuanced representation of mood, capturing the diversity of emotional experiences with a higher accuracy:

* **High Energy, Low Stress (Exuberance):** Represents a lively and happy emotional state. This mood is associated with feelings of excitement and enthusiasm, often elicited by upbeat and fast-paced music.
* **High Energy, High Stress (Anxious/Frantic):** Reflects an intense and tense state, characterized by feelings of anxiety or agitation. Music in this category often has a fast tempo but conveys tension through dissonance or abrupt changes.
* **Low Energy, Low Stress (Contentment):** Denotes a relaxed and pleasant emotional state. Music in this category is typically calm, slow, and soothing, evoking feelings of peace and tranquillity.
* **Low Energy, High Stress (Depression):** Represents a state of sadness and tension. Music that falls into this category is often slow, melancholic, and minor-key, conveying feelings of sorrow or fatigue.

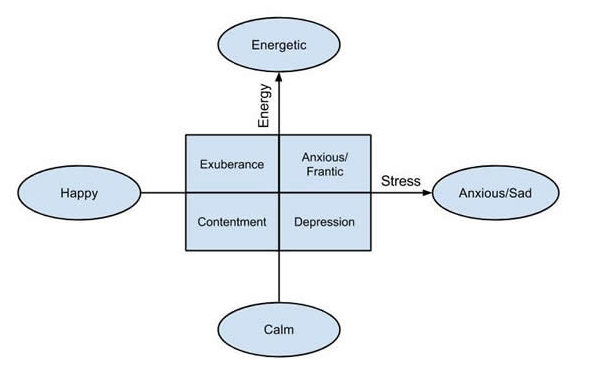


Figure 2: Thayer’s mood model and emotion categories [37].

**2.1.4 Machine Learning and AI in Music Recommendation**

The utilization of machine learning and artificial intelligence (AI) has fundamentally transformed the process of recommending music to users, as it has facilitated the creation of personalized and adaptable recommendation algorithms. These advances enable the creation of sophisticated systems that can analyze user behavior, preferences, and contextual data to provide tailored music recommendations, which develop over time to better correspond with individual preferences.

#### Music Recommendation Approaches

Similarly to recommendation systems in other domains, such as e-commerce or video streaming platforms, music recommendations systems have also traditionally leveraged both collaborative filtering and content-based techniques. These methods, which respectively rely on user behavior patterns and item attributes, have been widely employed to generate personalized music recommendations, and in some case combining these approaches has proven effective in enhancing recommendation accuracy and diversity.

1. **Collaborative Filtering**: Collaborative filtering is a methodology employed in music recommendation systems to generate predictions about user’s preferences by analyzing similarities of collective user interactions [26]. The system identifies patterns in users habits and behaviors, such as ratings and listening history, to recommend potential matches between existing and undiscovered tracks across different styles and genres [27]. Some limitations of collaborative filtering include [19]:

* Data sparsity: When there are a lot of users and items, some items receive very few ratings, making predictions less accurate.
* Long tail issue (popularity bias); Well-like items tend to be suggested more often, whereas lesser-known ones are seldom recommended.
* Cold start dilemma: Newly added items and users encounter challenges due to insufficient data for creating dependable recommendations.

1. **Content-Based Filtering**: Content-based filtering is a recommendation approach that leverages item attributes, such as genre, rhythm, and audio features. By analyzing the song’s content, content-base filtering suggests new music that shares similar characteristics with the user’s known preferences [28]. However, this approach has its own set of drawbacks [19]:

* User preference modeling: Difficulty to accurately understand what users like just by looking at similar content.
* Feature extraction: Extracting meaningful high-level descriptors of consistent type and quality is complex.
* Repetitive recommendations: Recommends items lack of novelty and diversity as recommendation are similar to previous songs.

1. **Hybrid Approaches**: The hybrid approach in music recommendation combines collaborative filtering and content-based filtering to leverage the strengths of both techniques while addressing their respective limitations. For instance a hybrid system could suggest music that aligns with a user's current preferences through content-base filtering, while employing collaborative filtering to introduce new and diverse options based on similar users' tastes. While hybrid systems may be more challenging to implement and demand more resources, they provide a more personalized and robust recommendation experience by considering both item attributes and user behavior [19].

#### Deep Learning in Music Recommendation

Deep learning models, such as Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have shown significant capabilities in improving music recommendation systems. These models can identify and learn complex patterns and correlation between users and data, by automatically categorizing a amalgamation of data types by capturing similarities between them [29]. By harnessing deep learning, these systems provide more personalized and context-aware music recommendations as it is able to handle large datasets and multiple modalities and learn intricate feature representations accurately [28].

#### AI and Context-Aware Recommendations

Artificial intelligence (AI) has enabled the development of context-aware music recommendation, by considering additional factors such as location, user activity, and time of day or season. This allows the recommendation system to provide music that aligns with the user's current needs and preferences, enhancing the overall user experience. For example, Kaminskas and Ricci [19] discussed the challenges and opportunities in contextual music information retrieval, highlighting how context-aware recommendations can enhance user satisfaction by enabling sophisticated and dynamic music recommendations that align with the user's emotional transitions and situational contexts​.

In conclusion, the integration of machine learning and artificial intelligence has paved the way for a new era of music recommendation, characterized by personalized, adaptive, and contextually relevant suggestions. Thus, future research should focus on exploring new ways to incorporate contextual and emotional data into the recommendation process, by fully exploring not onlycontextual parameters but also the relations between music and certain contextual conditions.

**2.2 Related Works**

This section provides an in-depth analysis of existing papers and systems, related to my proposed project. By analyzing the features, functionalities, and user experiences of these music recommendation systems, this comprehensive examination can identify the strengths and weakness of current systems and propose a relevant solution to fill the gaps. The selected papers for review are:

1. "Emotion Based Music Playlist Recommendation System using Interactive Chatbot," by Nair, A., Pillai, S., Nair, G. S., & Anjali, T. [4]
2. "Emotune: Emotion And Gender Aware Music Generation Chatbot," by A. P. Rao, Nithasha, P. R., Prathik S., and R. Rao. [29] [30]
3. "Emousic: Emotion and Activity-Based Music Player Using Machine Learning," by P. Sarda, S. Halasawade, A. Padmawar, and J. Aghav. [31]
4. "Induced Emotion-Based Music Recommendation through Reinforcement Learning," by R. De Prisco, A. Guarino, D. Malandrino, and R. Zaccagnino. [28]
5. “Smart music player integrating facial emotion recognition and music mood recommendation,” by S. Gilda, H. Zafar, C. Soni, and K. Waghurdekar. [32]

**2.2.1 "Emotion Based Music Playlist Recommendation System using Interactive Chatbot"**

**Summary**

Nair et al. (2021) developed an emotion-based music playlist recommendation system that utilizes an interactive chatbot to enhance user engagement and personalization. The system leverages natural language processing (NLP) techniques to analyze user input and determine their emotional state, subsequently recommending music that aligns with the identified emotions. This approach aims to provide a more personalized music experience by considering the user's current mood and preferences.

**Features and Functionality**

* System Platform: Mobile Application
  + The system is designed as a mobile application, making it accessible and convenient for users to interact with the chatbot and receive music recommendations on the go.
* Multi-modal Input: No
  + The system does not utilize multi-modal input methods such as voice or image recognition. Instead, it relies solely on text input for emotion detection.
* Text Input: Yes
  + Users interact with the chatbot via text input. The chatbot asks general questions related to various topics such as sports and weather to gauge the user's emotional state.
* User Emotion Recognition: Text
  + The system analyzes the text input provided by users to recognize their emotional state. This is achieved through sentiment analysis, which classifies responses as positive, negative, or neutral.
* Emotion Model: Polarity score [-1.0, 1.0]
  + The sentiment analysis model assigns a polarity score to each user response, with scores ranging from -1.0 (negative) to 1.0 (positive). The overall sentiment is determined based on the cumulative score of the responses.
* Techniques for Recognizing Emotion in Music: Audio
  + The system uses audio analysis to determine the emotional content of music tracks. Key attributes such as energy, valence, danceability, and acousticness are analyzed to classify the mood of the songs.
* Music Database: Spotify Web API
  + The system utilizes the Spotify Web API to access a vast database of music tracks. This allows for diverse and up-to-date music recommendations based on the user's emotional state.
* Machine Learning Models: Bidirectional LSTM
  + Three models were considered for sentiment analysis: LSTM, Bidirectional LSTM, and 1-D Convolutional Neural Network (CNN). The Bidirectional LSTM model was selected for its superior performance, achieving an accuracy of 79.29%.
* ML Model Accuracy: 79.29%
  + The Bidirectional LSTM model used for sentiment analysis achieved an accuracy of 79.29%, making it the most effective model for the system's needs.
* Recommendation Methods: Collaborative, AI, ML
  + The system combines collaborative filtering, artificial intelligence, and machine learning techniques to generate personalized music recommendations. This hybrid approach ensures that recommendations are tailored to the user's current emotional state and listening history.
* Chatbot Interaction: Yes
  + The interactive chatbot is a core component of the system. It engages users in conversation to determine their emotional state and provides personalized music recommendations based on the analyzed sentiment.
* Adaptive Player: Yes
  + The system features an adaptive music player that adjusts the playlist in real-time based on the user's mood. This is done through the users existing Spotify account data.

**Analysis**

The system's integration of a chatbot for emotion recognition and music recommendation is a significant strength, as it enhances user interaction and personalization. The use of a polarity score model for emotion detection, while effective, may not capture the full spectrum of human emotions. The reliance on text input alone could limit the accuracy of emotion detection, as it does not consider other modalities such as voice or facial expressions. However, the use of the Spotify Web API ensures a diverse music database, and the hybrid recommendation approach improves the quality and relevance of the suggestions.

**2.2.2 "Emotune: Emotion And Gender Aware Music Generation Chatbot"**

**Summary**

Rao et al. (2024) presented Emotune, an advanced music generation chatbot designed to be both emotion and gender-aware. This system employs a multi-modal approach, incorporating text, audio, and facial expressions to provide a comprehensive understanding of user emotions. The objective is to generate music that not only matches the emotional state of the user but also considers gender-specific preferences, thereby enhancing user engagement and personalization.

**Features and Functionality**

* System Platform: Website
  + Emotune is accessible through a web-based platform, allowing users to interact with the system and receive music recommendations from any device with internet access.
* Multi-modal Input: Yes
  + The system supports multiple input modalities, including text, audio, and image, which enhances the accuracy of emotion detection by analyzing various types of user data.
* Facial Expression Recognition: Yes
  + Emotune uses facial expression recognition technology to detect emotions based on the user's facial cues, adding another layer of emotional context.
* User Emotion Recognition: Text, Audio, Facial Expression
  + The combination of text, audio, and facial expression analysis allows for a robust and comprehensive assessment of the user's emotions.
* Emotion Model: Emotional spectrum (four primary types: neutral, happy, sad, and angry)
  + Emotions are categorized into four primary types: neutral, happy, sad, and angry, enabling the system to tailor music recommendations more effectively.
* Techniques for Recognizing Emotion in Music: Multimodal
  + Emotune employs a multi-modal approach to recognize emotions in music, integrating data from various input sources to provide accurate and personalized recommendations.
* Music Database: In-built, admin managed
  + The system uses an in-built music database managed by administrators, ensuring that the music selection is curated and relevant.
* Machine Learning Models: LSTM
  + Long Short-Term Memory (LSTM) models are used for processing sequential data and understanding the temporal dependencies in user inputs, which is crucial for accurate emotion detection.
* Recommendation Methods: Hybrid, AI, ML
  + The system utilizes a hybrid approach that combines collaborative filtering, artificial intelligence, and machine learning to generate music recommendations that align with the user's emotions and gender preferences.
* Chatbot Interaction: Yes
  + The interactive chatbot engages users in conversation to gather emotional and contextual data, enhancing the personalization of music recommendations.
* Adaptive Player: Yes
  + Emotune features an adaptive music player that adjusts the playlist based on real-time emotion detection, ensuring the music remains relevant to the user's current mood.
* Feedback Mechanism: Yes
  + Users can provide feedback on the recommended music, allowing the system to refine its algorithms and improve future recommendations.
* Social Sharing and Collaborative Playlists: Yes
  + The system supports social sharing and the creation of collaborative playlists, enabling users to import their own song into the application.

**Analysis**

Emotune's strength lies in its comprehensive multi-modal approach to emotion detection, which allows for a more nuanced and accurate understanding of user emotions. The integration of text, audio, and facial expression analysis provides a robust dataset for the system to work with, ensuring that music recommendations are highly personalized. The inclusion of gender-specific preferences adds an additional layer of customization, making the system more appealing to a diverse user base.

The use of LSTM models for emotion recognition is a significant advantage, as these models are well-suited for handling sequential data and capturing temporal dependencies. This enhances the accuracy of emotion detection and ensures that the music recommendations are aligned with the user's current emotional state. However, the reliance on an in-built music database managed by administrators could limit the diversity of the music recommendations compared to systems that use external APIs like Spotify. Additionally, while the system supports social sharing and collaborative playlists, the extent to which these features are integrated and utilized by users remains to be seen.

**2.2.3 "Emousic: Emotion and Activity-Based Music Player Using Machine Learning"**

**Summary**

Sarda et al. (2019) developed Emousic, an innovative music player that leverages machine learning to recommend music based on both the user's emotional state and physical activity. By integrating emotion detection and activity monitoring, Emousic aims to enhance the user's music listening experience by aligning song choices with their current mood and activity level.

**Features and Functionality**

* System Platform: Mobile Application
  + Emousic is designed as a mobile application, allowing users to interact with the system and receive music recommendations directly on their mobile devices.
* Multi-modal Input: Yes
  + The system supports multiple input modalities, including audio and sensor data, to accurately detect the user's emotional state and activity level.
* Image Input: Yes
  + Emousic can analyse images, particularly facial expressions, to further refine the understanding of the user's emotional state.
* Facial Expression Recognition: Yes
  + The system uses facial expression recognition technology to detect emotions based on the user's facial cues, enhancing the accuracy of emotion detection.
* Sensors: Yes
  + Emousic integrates sensor data to monitor the user's physical activity, such as walking, running, or sitting, to tailor music recommendations accordingly.
* User Emotion Recognition: Audio
  + The system primarily uses audio analysis to recognize the user's emotional state, utilizing features such as tone and pitch from speech or music.
* Techniques for Recognizing Emotion in Music: Manually tagged
  + Emousic relies on manually tagged music tracks, where each song is pre-labeled with its associated emotion, facilitating accurate music recommendations based on the detected user emotions.
* Music Database: In-built, 101 songs
  + The system uses an in-built music database containing 101 songs, ensuring that the music selection is curated and relevant to the user's emotional and activity states.
* Machine Learning Models: Random Forest
  + The system employs Random Forest models for both emotion and activity recognition, leveraging the model's robustness and accuracy in handling complex datasets.
* ML Model Accuracy: 96.33%
  + The Random Forest model used for emotion and activity recognition achieved an impressive accuracy of 96.33%, ensuring reliable detection and recommendations.
* Recommendation Methods: Context-based, AI, ML
  + Emousic uses a context-based recommendation approach, integrating artificial intelligence and machine learning techniques to generate music recommendations that align with the user's current emotional and activity state.
* Feedback Mechanism: Yes
  + Users can provide feedback on the recommended music, allowing the system to refine its algorithms and improve future recommendations.

**Analysis**

Emousic stands out due to its integration of both emotion detection and activity monitoring, providing a holistic approach to music recommendation. By considering both emotional and physical states, the system offers a more personalized and context-aware music experience. The use of Random Forest models for emotion and activity recognition is a significant strength, given the model's ability to handle large and complex datasets with high accuracy. The reported accuracy of 96.33% highlights the system's reliability in detecting user emotions and activities. However, the system's reliance on a manually tagged music database of only 101 songs may limit the diversity and freshness of the music recommendations. Expanding the music database and incorporating external music APIs could enhance the system's ability to provide varied and up-to-date recommendations.

**2.2.4 "Induced Emotion-Based Music Recommendation through Reinforcement Learning"**

**Summary**

De Prisco et al. (2022) introduced Moodify, an emotion-based music recommendation system that utilizes reinforcement learning to tailor music suggestions based on the user's induced emotions. By employing reinforcement learning techniques, Moodify aims to continuously learn and adapt to the user's preferences, providing more accurate and satisfying music recommendations over time.

**Features and Functionality**

* System Platform: Website
  + Moodify is accessible through a web-based platform, enabling users to interact with the system and receive music recommendations from any internet-connected device.
* Multi-modal Input: No
  + The system does not support multi-modal inputs such as voice or image recognition. It primarily relies on user interactions and feedback to understand preferences, like a standard mobile application.
* User Emotion Recognition: Not applicable
  + Unlike other systems, Moodify does not directly recognize user emotions through sensors or text analysis. Instead, it uses user feedback and interaction data to infer emotional preferences indirectly.
* Emotion Model: Circumplex model (four groups: Pleasant-high, Pleasant-low, Unpleasant-high, and Unpleasant-low)
  + The system categorizes emotions using the Circumplex model, which divides emotions into four groups based on their valence (pleasant-unpleasant) and arousal (high-low) dimensions.
* Techniques for Recognizing Emotion in Music: Audio
  + Moodify uses audio analysis to determine the emotional content of music tracks. Features such as tempo, rhythm, and melody are analyzed to classify the mood of the songs.
* Music Database: Spotify Web API
  + The system utilizes the Spotify Web API to access a vast database of music tracks, ensuring a diverse and up-to-date selection for recommendations.
* Machine Learning Models: Reinforcement Learning, Go-Explore
  + Moodify employs reinforcement learning techniques, specifically the Go-Explore algorithm, to continuously improve its music recommendations based on user feedback and interaction patterns.
* Recommendation Methods: Hybrid, AI, ML
  + The system combines hybrid recommendation methods, artificial intelligence, and machine learning to generate music recommendations that align with the user's inferred emotional preferences.
* Critiquing for Refining Recommendations: Yes
  + Moodify includes a critiquing mechanism that allows users to provide feedback on the recommended music. This feedback is used to refine the recommendation algorithms and improve future suggestions.
* Feedback Mechanism: Yes
  + Users can provide explicit feedback on the music recommendations, helping the system learn and adapt to their preferences more effectively.

**Analysis**

Moodify's use of reinforcement learning is a key strength, as it allows the system to continuously learn from user interactions and adapt its recommendations over time. This dynamic approach ensures that the system remains relevant and satisfying for users, as it can quickly adjust to changes in their preferences. The reliance on the Circumplex model for emotion categorization provides a structured framework for understanding and predicting user preferences based on emotional dimensions. This model, combined with the use of the Spotify Web API, ensures a diverse and accurate selection of music tracks. However, the absence of direct emotion recognition through sensors or text analysis could be seen as a limitation, as it relies heavily on user feedback to infer emotional states. Integrating multi-modal inputs in the future could enhance the system's ability to understand and respond to user emotions more accurately.

**2.2.5 “Smart music player integrating facial emotion recognition and music mood recommendation”**

**Summary**

Gilda et al. (2017) developed a smart music player that integrates facial emotion recognition with music mood recommendation to enhance the user's listening experience. This system utilizes advanced computer vision techniques to detect the user's emotional state through facial expressions and then recommends music that aligns with the detected mood. The primary aim is to create a seamless and personalized music experience by leveraging real-time emotion detection.

**Features and Functionality**

* System Platform: Mobile Application
  + The smart music player is designed as a mobile application, providing users with a portable and convenient platform to interact with the system and receive music recommendations.
* Multi-modal Input: Yes
  + The system supports multiple input modalities, including image and sensor data, to accurately detect the user's emotional state.
* User Emotion Recognition: Facial Expression
  + The primary method for recognizing user emotions is through the analysis of facial expressions captured by the device's camera.
* Emotion Model: Russell 2-D Valence-Arousal Model and Geneva Emotion Wheel
  + The system utilizes the Russell 2-D Valence-Arousal Model and the Geneva Emotion Wheel to classify and understand user emotions, providing a comprehensive framework for emotion detection.
* Techniques for Recognizing Emotion in Music: Audio
  + Music tracks are analyzed for their emotional content using audio features such as tempo, rhythm, and melody to classify the mood of the songs.
* Music Database: In-built, 390 songs
  + The system uses an in-built music database containing 390 songs, curated to cover a wide range of emotional states and musical genres.
* Machine Learning Models: Convolutional Neural Networks (CNN)
  + Convolutional Neural Networks (CNN) are employed for facial expression recognition, leveraging the model's ability to handle image data and detect complex patterns in facial expressions.
* ML Model Accuracy: 90.23%
  + The CNN model used for facial emotion recognition achieves an accuracy of 90.23%, ensuring reliable and accurate emotion detection.
* Recommendation Methods: Hybrid, Context-based
  + The system uses a hybrid approach that combines collaborative filtering, context-based recommendations, and machine learning techniques to generate personalized music recommendations.
* Adaptive Player: Yes
  + The smart music player features an adaptive player that adjusts the playlist in real-time using the Stochastic Gradient Descent (SGD) algorithm to personalize the labels of songs according to the user’s preferences.
* Critiquing for Refining Recommendations: Yes
  + Users can provide feedback on the recommended music, allowing the system to refine its algorithms and improve future recommendations.
* Feedback Mechanism: Yes
  + The application includes a feedback mechanism that lets users rate and review the music recommendations, aiding in the continuous improvement of the recommendation system.

**Analysis**

The smart music player's integration of facial emotion recognition is a significant advantage, as it provides real-time and accurate detection of user emotions. This approach ensures that music recommendations are highly personalized and aligned with the user's current emotional state. The use of CNNs for facial expression recognition further enhances the system's reliability, given the model's proven effectiveness in handling image data. The dual use of the Russell 2-D Valence-Arousal Model and the Geneva Emotion Wheel offers a comprehensive framework for understanding and classifying emotions, which enhances the accuracy of emotion detection and music recommendation. Additionally, the inclusion of an adaptive player and feedback mechanism ensures that the system remains responsive to user preferences and continuously improves over time. However, the reliance on an in-built music database could limit the diversity and freshness of music recommendations. Expanding the music database or integrating external music APIs could enhance the system's ability to provide a more varied and up-to-date selection of music.

**2.3 Findings and Results**

This section presents the findings and results from the comparative analysis of various emotion-based music recommendation systems. The analysis aims to identify the strengths and limitations of existing systems, providing a comprehensive understanding of their features and functionalities. By examining key features like system platforms, input methods, emotion recognition technologies and machine learning models we can assess how well these systems meet user preferences and requirements.

The analysis reveals distinctions in how each system handles emotion detection and music suggestions. Some systems use types of input including text, audio and facial cues for accurate emotion recognition; others rely on just one type like text or audio. The choice of emotion models and machine learning techniques varies well impacting the quality and precision of recommendations. Furthermore features like players, feedback options and social sharing functions enhance the overall user experience by highlighting areas for potential enhancement, in future iterations.

The following table summarizes the main features and functionalities of the five related systems analysed:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Feature | Emotion-Based System | Emotune | Emousic | Moodify | Smart Music Player |
| System Platform | Mobile Application | Website | Mobile Application | Website | Mobile Application |
| Multi-modal Input | No | Yes | Yes | No | Yes |
| Text Input | Yes | Yes | No | No | No |
| Voice Input | No | Yes | No | No | No |
| Image Input | No | Yes | Yes | No | Yes |
| Facial Expression Recognition | No | Yes | Yes | No | Yes |
| Sensors | No | No | Yes | No | Yes |
| User Emotion Recognition | Text | Text, Audio, Facial Expression | Audio | Not applicable | Facial Expression |
| Emotion Model | Polarity score  [-1.0, 1.0] | Emotional spectrum (four primary types: neutral, happy, sad, and angry) | Not specified | Circumplex model (four groups: Pleasant-high, Pleasant-low, Unpleasant-high, and Unpleasant-low) | Russell 2-D Valence-Arousal Model and Geneva Emotion Wheel |
| Techniques for Recognizing Emotion in Music | Audio | Multimodal | Manually tagged | Audio | Audio |
| Music Database | Spotify Web API | In-built, admin managed | In-built, 101 songs | Spotify Web API | In-built, 390 songs |
| Machine Learning Models | Bidirectional LSTM | LSTM | Random Forest | Reinforcement Learning, Go-Explore | CNN |
| ML Model Accuracy | 79.29% | Not specified | 96.33% | Not specified | 90.23% |
| Recommendation Methods | Collaborative, AI, ML | Hybrid, AI, ML | Context-based, AI, ML | Hybrid, AI, ML | Hybrid, Context-based |
| Chatbot Interaction | Yes | Yes | No | No | No |
| Adaptive Player | Yes | Yes | No | Yes | Yes |
| Critiquing for Refining Recommendations | No | No | No | Yes | Yes |
| Feedback Mechanism | No | Yes | Yes | Yes | Yes |
| Social Sharing and Collaborative Playlists | No | Yes | No | No | No |

Table 4: Comparison of Key Features and Functionalities of Current Systems

1. **Emotion**-**Based System**:

Nair, A., Pillai, S., Nair, G. S., & Anjali, T. "Emotion Based Music Playlist Recommendation System using Interactive Chatbot," Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES-2021), pp. 1767-1772, 2021. DOI: 10.1109/ICCES51350.2021.9489138. Available: https://ieeexplore.ieee.org/document/9489138.

1. **Emotune**:

A. P. Rao, Nithasha, P. R., Prathik S., and R. Rao, "Emotune: Emotion And Gender Aware Music Generation Chatbot," International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), vol. 13, no. 4, pp. 1044-1059, Apr. 2024, doi: 10.17148/IJARCCE.2024.134152. Available: IJARCCE.

A. P. Rao, Nithasha, P. R., Prathik S., and R. Rao, "EMOTUNE", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.12, Issue 1, pp.b434-b445, January 2024, Available at :http://www.ijcrt.org/papers/IJCRT2401181.pdf

1. **Emousic**:

P. Sarda, S. Halasawade, A. Padmawar, and J. Aghav, "Emousic: Emotion and Activity-Based Music Player Using Machine Learning," in Advances in Computer Communication and Computational Sciences, S. K. Bhatia et al., Eds. Springer, Singapore, 2019, pp. 179-188, doi: 10.1007/978-981-13-6861-5\_16. Available: SpringerLink.

1. **Moodify**:

R. De Prisco, A. Guarino, D. Malandrino, and R. Zaccagnino, "Induced Emotion-Based Music Recommendation through Reinforcement Learning," Applied Sciences, vol. 12, no. 21, p. 11209, Nov. 2022, doi: 10.3390/app122111209. Available: MDPI.

1. **Smart Music Player**:

S. Gilda, H. Zafar, C. Soni, and K. Waghurdekar, “Smart music player integrating facial emotion recognition and music mood recommendation,” 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Mar. 2017, doi: 10.1109/wispnet.2017.8299738.

**2.3.1 Comparative Analysis of Existing Systems**

**Emotion-Based Music Playlist Recommendation System using Interactive Chatbot**

Strengths:

* **Text Input and Chatbot Interaction**: The integration of a chatbot for user interaction makes the system user-friendly and interactive. The chatbot can gather detailed emotional feedback through text, enhancing the personalization of music recommendations.
* **Adaptive Player**: The system features an adaptive player that adjusts playlists in real-time based on the user’s mood, providing a highly personalized listening experience.
* **Machine Learning Models**: Utilizes Bidirectional LSTM for sentiment analysis, achieving a respectable accuracy of 79.29%. This model effectively captures the context of user inputs, leading to more accurate emotion detection.
* **Music Database**: Leverages the Spotify Web API, ensuring a vast and diverse selection of music tracks that are regularly updated.

Weaknesses:

* **Limited Input Modalities**: The system relies solely on text input for emotion detection, missing out on additional emotional cues that could be gathered from voice or facial expressions.
* **Lack of Feedback Mechanism**: The absence of a feedback mechanism limits the system's ability to learn and adapt from user preferences over time.
* **No Social Sharing Features**: The system does not support social sharing or collaborative playlists, which could enhance user engagement and discovery of new music.

**Emotune: Emotion and Gender Aware Music Generation Chatbot**

Strengths:

* **Multi-modal Input**: Emotune supports text, audio, and facial expression inputs, allowing for a robust and comprehensive assessment of user emotions.
* **Gender-Aware**: The system considers gender-specific preferences, adding an extra layer of personalization to music recommendations.
* **Hybrid Recommendation Methods**: Utilizes a combination of collaborative filtering, AI, and ML, which enhances the diversity and accuracy of recommendations.
* **Social Sharing and Collaborative Playlists**: Supports social sharing and the creation of collaborative playlists, fostering community engagement and music discovery.

Weaknesses:

* **In-built Music Database**: Relies on an in-built music database managed by administrators, which may limit the diversity and freshness of music recommendations compared to systems that use external APIs like Spotify.
* **Unspecified Model Accuracy**: The accuracy of the LSTM models used for emotion detection is not specified, making it difficult to assess the system's reliability.

**Emousic: Emotion and Activity-Based Music Player Using Machine Learning**

Strengths:

* **Multi-modal Input**: Emousic integrates sensor data and facial expressions for emotion detection, providing a holistic approach to understanding user emotions and activities.
* **High Model Accuracy**: The use of Random Forest models for emotion and activity recognition achieves a high accuracy of 96.33%, ensuring reliable emotion detection.
* **Context-based Recommendations**: Tailors music recommendations based on the user's current emotional and physical state, providing a highly personalized listening experience.

Weaknesses:

* **Limited Music Database**: The system uses a manually tagged in-built music database with only 101 songs, which limits the diversity and novelty of music recommendations.
* **No Chatbot Interaction**: Lacks a chatbot for user interaction, which could enhance user engagement and provide more detailed emotional feedback.
* **No Social Sharing Features**: Does not support social sharing or collaborative playlists, which could limit user engagement and music discovery.

**Moodify: Induced Emotion-Based Music Recommendation through Reinforcement Learning**

Strengths:

* **Reinforcement Learning**: Utilizes reinforcement learning techniques, specifically the Go-Explore algorithm, to continuously learn and adapt to user preferences, ensuring that music recommendations improve over time.
* **Feedback Mechanism**: Emphasizes user feedback and critiquing, allowing the system to refine its recommendations based on direct user input.
* **Context-aware Recommendations**: Provides context-aware recommendations by analysing user interactions and feedback, which helps in understanding user preferences better.

Weaknesses:

* **Lack of Multi-modal Input**: The system does not utilize multi-modal inputs like voice or facial recognition, which could limit the accuracy of emotion detection.
* **No Direct Emotion Recognition**: Relies on indirect emotion recognition through user feedback rather than real-time emotion detection, which may not capture immediate emotional changes.
* **Music Database**: Uses Spotify Web API, which is beneficial for music diversity but might lack curated personalization compared to in-built databases managed by administrators.

**Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation**

Strengths:

* **Facial Emotion Recognition**: Integrates facial emotion recognition technology, which allows for accurate real-time detection of user emotions based on facial expressions.
* **High Model Accuracy**: Uses CNN models for facial emotion recognition, achieving a high accuracy of 90.23%, ensuring reliable emotion detection.
* **Comprehensive Emotion Models**: Utilizes the Russell 2-D Valence-Arousal Model and Geneva Emotion Wheel, providing a detailed framework for emotion classification.
* **Adaptive Player**: Features an adaptive player that adjusts the playlist in real-time based on the detected emotional state of the user.

Weaknesses:

* **Limited Music Database**: Relies on an in-built music database with 390 songs, which may limit the diversity and novelty of music recommendations.
* **No Chatbot Interaction**: Does not include a chatbot for user interaction, which could enhance user engagement and provide more detailed emotional feedback.
* **Lack of Multi-modal Input**: Primarily uses facial expressions for emotion detection, missing out on additional emotional cues from voice or text.

**Summary**

By comparing the strengths and weaknesses of these existing systems, it is evident that there are several areas where the proposed solution can provide significant improvements. The integration of multi-modal input methods, detailed feedback mechanisms, adaptive features, and social sharing capabilities will address the limitations of current systems, offering a more comprehensive and user-centric music recommendation experience. The proposed system will leverage advanced AI and ML techniques coupled with user feedback, to continuously learn and adapt to user preferences, ensuring that music recommendations are always relevant and personalized.

**2.4 Proposed Solution**

This section outlines the proposed system, " VibeSync: An Emotion-Based Recommendation System with Integrated Chatbot", designed to address the limitations identified in the existing systems. VibeSync aims to provide a comprehensive and user-centric music recommendation experience by leveraging advanced AI and machine learning techniques, multi-modal input methods, and detailed feedback mechanisms.

**Overview of VibeSync**

VibeSync is envisioned as a standalone mobile application with potential integration with existing music platforms like Spotify. The system's primary goal is to enhance the user's music listening experience by accurately detecting and responding to their emotional states through a combination of text, audio, and potentially facial expression inputs. By incorporating adaptive features, context-aware recommendations, and social sharing capabilities, VibeSync aims to offer a more personalized and engaging music recommendation service.

**Key Features and Enhancements**

1. System Integration
   * **Standalone Application**: VibeSync will function as a standalone mobile application, providing users with a seamless and intuitive interface for music recommendations.
   * **Integration with Music Platforms**: The application will integrate with existing music platforms like Spotify, ensuring access to a vast and diverse music database. This integration will address the limitations of in-built music databases by offering a continuously updated and extensive selection of music tracks.
2. User Input Methods
   * **Text Input**: Users can interact with the integrated chatbot through text input, allowing the system to gather detailed emotional feedback and preferences.
3. Emotion Recognition
   * **Emotion Recognition from Text**: The system will use transfer learning to fine-tune a pre-trained model for emotion detection based on text. Using a pre-trained model allow the system to take advantage of the source database, while still creating a specified model for the system.
4. Recommendation Methods
   * **Content-based Filtering**: VibeSync will use mainly utilize content-based filtering to recommend songs based on weighted emotion labels, with the detected user emotion. However, the user will still be able to get recommendations based on their music preferences as they are able to add their own songs into the system.
5. Interaction Modes
   * **Integrated Chatbot**: The chatbot will facilitate user interaction, gathering emotional feedback and providing music recommendations based on detected emotions. The chatbot will engage users in natural conversations, enhancing the personalization and relevance of recommendations.
   * **Website Application Interface**: The user-friendly mobile application interface will allow users to easily navigate and interact with the system, ensuring a smooth and enjoyable user experience.
6. Adaptive Features
   * **Adaptive Player**: The adaptive player will adjust playlists in real-time based on the user's detected emotions. By continuously monitoring user feedback and emotional state, the player will ensure that the music aligns with the user's current mood and preferences, enhancing the overall listening experience.
7. Feedback Mechanisms
   * **Explicit Feedback**: Users will be able to provide explicit feedback, on the system and the song classification and recommendation. This feedback will be used to improve the system experience, refine the recommendation algorithms and fine-tune future suggestions.
8. Personalized Playlist Generation
   * **Emotional State and Preferences**: The system will generate personalized playlists based on the user's emotional state and preferences. By analysing real-time emotion feedback, VibeSync will curate playlists that match or influence the user's mood.
   * **Save and Edit Playlists**: Users will have the option to save and edit playlists, providing greater control over their music experience and allowing for continuous customization.
   * **User-Generated Playlists**: VibeSync will enable users to enter their own playlists into the system. The system will classify the input songs and provide recommendations from the user-generated playlist, ensuring that personal favorites and curated lists are integrated into the recommendation process.
9. Context-aware Recommendations
   * **Adaptive to Mood Changes**: By adapting to the user’s current mood in real-time, VibeSync will ensure that the music recommendations remain relevant and enhance the user's current experience.
10. User-friendly Interface
    * **Intuitive Design**: VibeSync will feature an intuitive and easy-to-use interface, ensuring that users can navigate the application effortlessly. The design will prioritize user experience and accessibility.
    * **Visualizations**: The application will provide visualizations of mood transitions and playlist adjustments, helping users understand how their emotions influence music recommendations.
11. Additional Unique Features
    * **Mood-based Exploration**: VibeSync will allow users to explore new music genres based on their emotional state. This feature will help users discover music that aligns with their mood and broadens their musical horizons.
    * **Weighted Emotion Song Tagging**: The application will use weighted emotion song tagging to provide more precise and personalized recommendations. This feature will ensure that the recommended songs closely match the user's emotional state and preferences.

**Addressing Limitations of Existing Systems**

The proposed features of VibeSync are designed to address the limitations identified in the existing systems. By incorporating multi-modal input methods, VibeSync will provide a more comprehensive understanding of user emotions, enhancing the accuracy of emotion detection. The integration of detailed feedback mechanisms, including both explicit and implicit feedback, will enable continuous learning and adaptation, improving the relevance and satisfaction of music recommendations.

VibeSync's adaptive player and context-aware recommendations will ensure that the music aligns with the user's current mood and situational context, offering a highly personalized and engaging listening experience. By leveraging advanced AI and machine learning techniques, VibeSync will continuously learn from user interactions and refine its recommendations, ensuring that the system remains relevant and satisfying over time. The proposed solution represents a comprehensive and user-centric approach to emotion-based music recommendation, addressing the gaps in existing systems and providing a more personalized and engaging music experience.

**Conclusion**

In conclusion, VibeSync aims to revolutionize emotion-based music recommendation by integrating advanced AI and machine learning techniques, multi-modal input methods, and detailed feedback mechanisms. By addressing the limitations of existing systems and incorporating innovative features, VibeSync will provide a comprehensive and user-centric music recommendation experience, enhancing user satisfaction and engagement. The proposed solution represents a significant step forward in the field of personalized music recommendation, offering a more accurate, adaptive, and enjoyable music listening experience for users.

# 3.0 Methodology

## 3.1 Tools

Here is a potential comprehensive list of the necessary tools for developing VibeSync.

|  |  |  |
| --- | --- | --- |
| No. | Tools | Purpose |
| 1 | Microsoft Word | Writing project documentation and maintaining activity logs. |
| 2 | Figma | Designing and prototyping the user interface for the website and chatbot. |
| 3 | Visual Studio Code | Integrated Development Environment (IDE) for coding, testing, and deployment. |
| 4 | React | Framework for building the dynamic, interactive website front-end. |
| 5 | HTML, CSS, JavaScript | Core technologies for front-end development, supported by frameworks like Bootstrap. |
| 6 | Python | Backend development, machine learning model development, and data processing. |
| 7 | Flask | Web framework for building REST API endpoints to connect the front-end with back-end services. |
| 8 | SQLite | Lightweight relational database for managing user data, song classifications, and preferences. |
| 9 | YouTube Data API, Spotify Web API | Accessing music data for playlists and song features. |
| 10 | Hugging Face | Using pre-trained models for emotion detection in user text inputs. |
| 11 | Postman | Testing API endpoints |
| 12 | TensorFlow, Keras | Building and fine-tuning machine learning models, including deep learning for song classification. |

Table 5: Tools and Their Usage in This Project

## 3.2 Data Collection and Processing

### 3.3.1 Music and Emotion Data

The primary objective of this data collection phase is to obtain a music dataset annotated with emotion labels to facilitate the classification of songs into weighted emotional categories.

**Data Source:**

Two datasets were obtained in this section.

1. **MER - Bimodal Dataset** [34]: This dataset contains 133 audio clips and lyrics, manually annotated into 4 quadrants, according to Russell's model. It includes a collection of songs, each annotated with valence and arousal scores derived from both audio and lyrics. These bimodal annotations provide a holistic view of emotional content in music, compared to other datasets that extracts the valence and arousal from just the lyrics or audio only. A python script was written to merge the two dataset by average valence and average arousal, and then fetch each songs’ unique features from the Spotify API. The features extracted include, Duration, Popularity, Danceability, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, and Tempo.
   * **Data Processing:** Data cleaning was performed to remove songs without Spotify features and duplicate entries, to maintain data quality and consistency.
2. **Moodify Database [**https://www.kaggle.com/datasets/abdullahorzan/moodify-dataset**]:** This dataset contains around 278 000 songs from Spotify with extracted features. The dataset also contains predicted emotion labels of Thayer’s emotion model using a LightGBM model created in Moodify.
   * **Data Processing:** The dataset was split into smaller, processable sets using stratified sampling to ensure that the smaller set accurately represents the distribution of the main set.

A screenshot of a computer

Description automatically generated

Figure 3: Snippet of the processed MER database.

### 3.2.2 Chatbot Text Data

To fine-tune the chatbot’s emotion detection capabilities, 8 publicly available text emotion datasets were utilized. The text samples include google lexicon, Reddit comments and post, Twitter post and movie scripts. These datasets contained annotations mapping textual inputs to specific emotion categories such as joy, sadness, anger, fear, and others.

**Data Source:**

1. GoEmotions Dataset [https://www.kaggle.com/datasets/debarshichanda/goemotions]
2. [https://github.com/SannketNikam/Emotion-Detection-in-Text]
3. [https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text]
4. [https://www.kaggle.com/datasets/simaanjali/emotion-analysis-based-on-text]
5. MELD Dataset [https://github.com/declare-lab/MELD?tab=readme-ov-file]
6. [https://www.kaggle.com/datasets/gargmanas/sentimental-analysis-for-tweets]
7. [https://www.kaggle.com/datasets/infamouscoder/depression-reddit-cleaned]
8. [https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch]

**Data Processing:**

* The text samples and emotion labels from all datasets were filtered to retain only categories relevant to the chatbot’s emotion scope:
  + happy, excited, energetic, fear, sad, depressed, calm, angry
* Text samples were cleaned and pre-processed by removing special characters, links, stop words, and irrelevant text.
* The cleaned datasets were then merged into one main dataset mapping a text string to an emotion.
* The dataset was then sampled to get an even distribution across all emotions. 30000 rows for each emotion were selected by random sampling for emotions exceeding the limit and duplication for emotions below the threshold.

A graph showing emotions and feelings

Description automatically generated

Figure 4: Dataset emotion distribution before processing.

**A graph of emotions in dataset

Description automatically generated**

Figure 5: Dataset emotion distribution after processing.

## 3.3 Detect Emotion of Songs with Weighted Labels

The methodology used to classify songs based on weighted emotion labels. The process involved data collection from multiple sources, preprocessing to extract relevant features, and training a machine learning model capable of predicting weighted emotion scores for new songs. By leveraging datasets with annotated emotional values and advanced classification algorithms, this system ensures accurate and nuanced emotion-based categorization of songs.

### 3.3.1 Valence Arousal Prediction

Valence-Arousal (V-A) prediction is a technique used in affective computing and music emotion recognition to model emotions on a two-dimensional space. This approach is based on the circumplex model, where [21]:

* **Valence** measures the positivity or negativity of an emotional state.
* **Arousal** reflects the intensity or energy level of the emotion.

The Valence-Arousal model is a foundational framework in emotion recognition, representing emotions on two dimensions: valence (positivity or negativity) and arousal (intensity). This study focuses on predicting these two dimensions to align musical features with emotional states. Spotify provides a variety of features, including a normalized valence score (0 to 1). However, to fit the Russell emotion circumplex model, valence values need to be remapped to span negative to positive ranges. For arousal prediction, Spotify song features, such as tempo, energy, and danceability were used to predict arousal values from the MER dataset’s average arousal scores.

The methodology for valence and arousal predictions involved:

* Model comparison to determine the best prediction model.
* Final prediction using the selected model.

#### Valence Prediction

*Objective*

To remap Spotify valence scores to a negative-positive range, aligning with the Russell model, using the MER dataset’s average valence scores as the target.

*Data and Features*

* **Input Variable:** Spotify valence scores (0 to 1).
* **Target Variable**: MER dataset’s average valence scores.
* **Data split**: Training (80%) and testing (20%).

*Model Comparison*

To identify the best approach for rescaling, several regression models were tested:

* Linear Regression
* SVR (RBF Kernel)
* SVR (Linear Kernel)
* Random Forest Regressor
* Gradient Boosting Regressor
* Huber Regressor
* TheilSen Regressor

*Evaluation Metrics*

* **R² Score**: Variance explained by the model.
* **RMSE (Root Mean Square Error)**: Magnitude of prediction error.
* **Cross-Validation (CV) Mean R² and Standard Deviation**: Measures model stability across folds.

*Results Model Comparison*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **R² Score** | **RMSE** | **CV Mean R²** | **CV Std R²** |
| **Linear Regression** | 0.136172 | 2.279128 | 0.105516 | 0.172641 |
| **TheilSen Regression** | 0.103296 | 2.322092 | 0.085891 | 0.105515 |
| **Huber Regression** | 0.127816 | 2.290124 | 0.080427 | 0.233852 |
| **SVR (Linear)** | 0.102900 | 2.322605 | 0.057789 | 0.289146 |
| **SVR (RBF)** | 0.247488 | 2.127216 | 0.004735 | 0.343952 |
| **Gradient Boosting** | 0.381417 | 1.928653 | -0.240211 | 0.264005 |
| **Random Forest** | 0.322545 | 2.018344 | -0.322268 | 0.154191 |

Table 5: Valence prediction model comparisons.

**Selected Model:** Based on these results, Linear Regression was selected for the final valence scaling due to its simplicity and reasonable performance. While Gradient Boosting showed a higher R², Linear Regression was preferred for its interpretability and lower computational complexity.

#### Arousal Prediction

*Objective*

To predict arousal values using Spotify features and the MER dataset’s average arousal as the target.

*Features Used*

The following Spotify features were used as predictors [38]:

* **Sp\_Duration\_ms**: Duration of the song in milliseconds.
* **Sp\_Danceability**: Measures how suitable a track is for dancing.
* **Sp\_Energy**: Perceived intensity and activity level of the track.
* **Sp\_Loudness**: Overall loudness of the song.
* **Sp\_Speechiness**: Presence of spoken words in the track.
* **Sp\_Acousticness**: Confidence level of whether the track is acoustic.
* **Sp\_Instrumentalness**: Confidence level of whether the track is instrumental.
* **Sp\_Liveness**: Presence of a live audience in the recording.
* **Sp\_Valence**: Positivity or negativity of the track.
* **Sp\_Tempo**: Overall tempo of the song.
* **Sp\_SpeechRate**: Speed of spoken lyrics or words.

*Data and Features*

* **Input Variable:** Spotify features
* **Target Variable**: MER dataset’s average arousal scores.
* **Data split**: Training (80%) and testing (20%).

*Model Comparison*

To identify the best approach for rescaling, several regression models were tested:

* Linear Regression
* Lasso Regression
* Ridge Regression
* Random Forest Regressor
* XGBoost Regressor
* CatBoost Regressor
* SVR

*Evaluation Metrics*

* **R² Score**: Variance explained by the model.
* **RMSE (Root Mean Square Error)**: Magnitude of prediction error.
* **Cross-Validation (CV) Mean R² and Standard Deviation**: Measures model stability across folds.

*Results Model Comparison*

*A group of colorful bars

Description automatically generated with medium confidence*

Figure 6: Arousal prediction model comparison results.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Position** | **R² Score** | **RMSE** | **CV Mean R²** | **CV Std** | **Overall Score** |
| **XGBoost** | 1 | 0.8163 | 1.0469 | 0.7987 | 0.0982 | 82.60 |
| **CatBoost** | 2 | 0.7887 | 1.1229 | 0.7701 | 0.0744 | 78.12 |
| **Random Forest** | 3 | 0.7647 | 1.1849 | 0.7393 | 0.1148 | 62.47 |
| **Lasso** | 4 | 0.7500 | 1.2215 | 0.7250 | 0.0576 | 56.50 |
| **Ridge** | 5 | 0.7291 | 1.2714 | 0.7045 | 0.0568 | 28.80 |
| **Linear Regression** | 6 | 0.7289 | 1.2718 | 0.7044 | 0.0579 | 27.44 |
| **SVR** | 7 | 0.6885 | 1.3633 | 0.6700 | 0.0427 | 16.77 |

Table 6: Arousal prediction model comparisons.

**Selected Model:** Based on the comparison, XGBoost was selected as the best-performing model due to its superior R² score (0.8163), lowest RMSE (1.0469), and high cross-validation stability.

#### Emotion Quadrant Classification

*Objective*  
To classify songs into emotional quadrants using the predicted valence and arousal values based on Russell’s Emotion Model.

*Data and Features*

* **Predicted Valence**: Obtained using the Linear Regression model selected during valence prediction.
* **Predicted Arousal**: Generated by the XGBoost model chosen during arousal prediction.

*Quadrant Assignment*

Songs were categorized into four quadrants based on their position in the valence-arousal space:

* **Quadrant 1**: High Arousal + Positive Valence (e.g., Excitement).
* **Quadrant 2**: High Arousal + Negative Valence (e.g., Anxiety).
* **Quadrant 3**: Low Arousal + Negative Valence (e.g., Sadness).
* **Quadrant 4**: Low Arousal + Positive Valence (e.g., Contentment).

A diagram of different emotions

Description automatically generated

Figure 7: Emotion Quadrants established in Russell’s circumplex model [39].

*Output*  
Each song was labelled with its corresponding emotion quadrant, enabling a comprehensive emotional analysis of the dataset. This classification provided interpretable results, mapping musical features to emotional states.

### 3.3.2 Weighted Emotion Classification

This phase builds upon the Valence-Arousal (V-A) predictions to provide a more detailed emotional classification for songs. While V-A prediction assigns songs to broad emotional quadrants, this step refines the emotional analysis by incorporating additional audio features and assigning weighted emotion labels based on linguistic evaluations. This approach creates a nuanced representation of each song’s emotional profile, essential for applications such as personalized recommendations or mood-based playlists.

#### Valence-Arousal Prediction

The initial step utilized the Valence-Arousal prediction model to classify songs in the Moodify Database into one of four quadrants:

* Quadrant 1: High Arousal + Positive Valence
* Quadrant 2: High Arousal + Negative Valence
* Quadrant 3: Low Arousal + Negative Valence
* Quadrant 4: Low Arousal + Positive Valence

Each quadrant broadly represents an emotional state. This classification formed the basis for further analysis, linking songs to general emotional categories.

#### Stratified Sampling

To ensure balanced representation, 1,000 songs were randomly selected from each quadrant. This stratified sampling approach ensured an equal distribution of emotional categories, maintaining dataset diversity while facilitating consistent downstream processing.

#### Audio Feature Extraction

To refine the emotional classification, audio features were extracted for the sampled songs.

*Retrieval and Preprocessing:*

Songs were downloaded and temporarily stored for analysis. Preprocessing ensured clean and standardized audio inputs for feature computation.

*Feature Extraction:*

Using Librosa, key features relevant to emotional characterization were extracted:

* Intensity: Measured using root mean square (RMS) energy, capturing the loudness and dynamics of a song.
* Timbre: Quantified using spectral centroid and bandwidth, representing the tonal color of the audio.
* Pitch: Derived from harmonic analysis, indicating the perceived frequency content.
* Rhythm: Determined by tempo and beat tracking.

*Data Management:*

Extracted features were structured into a dataset for analysis, and temporary files were deleted to optimize storage.

#### Linguistic Value Mapping

To facilitate interpretability, the numerical values of the extracted audio features were categorized into linguistic ranges such as "Very Low," "Low," "Medium," "High," and "Very High."

* Range Determination: Percentile thresholds were calculated for each feature across the dataset.
* Assignment: Each feature value was mapped to the corresponding linguistic category.

This mapping aligned raw numerical data with human-perceivable categories, enabling their use in emotion classification. Each linguistic feature was divided into bins representing different intensity levels. These bins served as the foundation for aligning song features with target emotion profiles:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Feature** | **Very Low** | **Low** | **Medium** | **High** | **Very High** |
| **Intensity** | <= 0.0938 | > 0.0938 and <= 0.1462 | > 0.1462 and <= 0.1955 | > 0.1955 and <= 0.2451 | > 0.2451 |
| **Timbre** | <= -6.8848 | > -6.8848 and <= -1.2937 | > -1.2937 and <= 2.4315 | > 2.4315 and <= 6.0919 | > 6.0919 |
| **Pitch** | <= 810.1318 | > 810.1318 and <= 1011.7241 | > 1011.7241 and <= 1195.7592 | > 1195.7592 and <= 1421.4669 | > 1421.4669 |
| **Rhythm** | <= 95.7031 | > 95.7031 and <= 112.3471 | > 112.3471 and <= 129.1992 | > 129.1992 and <= 143.5547 | > 143.5547 |

These ranges were determined using statistical analysis of the extracted features, ensuring that each feature aligns with the target emotion categories defined in the Emotion Table.

#### Emotion Classification

*Emotion Table*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Mood** | **Intensity** | **Timbre** | **Pitch** | **Rhythm** |
| **Happy** | Medium | Medium | Very High | Very High |
| **Excited** | High | Medium | High | High |
| **Energetic** | Very High | Medium | Medium | High |
| **Fear** | High | High | Very High | Very High |
| **Sad** | Medium | Very Low | Very Low | Low |
| **Depressed** | Low | Low | Low | Low |
| **Calm** | Very Low | Very Low | Medium | Very Low |
| **Angry** | Very High | Very High | Medium | Very High |

Table 7: Emotions classified according to musical components [37] [40].

The Emotion Table was developed as a systematic framework to map musical features such as intensity, timbre, pitch, and rhythm to specific emotional categories. This approach integrates insights from studies exploring how these auditory features correspond to emotional states in music and speech, providing a structured method for weighted emotion classification.

Intensity, often linked to the arousal dimension of emotions, plays a critical role in emotional representation. High-energy emotions such as happiness, excitement, and anger are associated with greater dynamics, reflecting higher intensity. Conversely, low-energy emotions like calmness and sadness exhibit softer dynamics and subdued loudness [37] [41]. For example, happiness tends to exhibit medium intensity, while calmness is characterized by very low intensity. Pitch is another universal indicator of emotional states. Higher pitches are often associated with positive emotions such as happiness and friendliness, while lower pitches convey more serious or aggressive emotions such as sadness or anger [42] [43]. These relationships have been observed consistently across both speech and music, highlighting the role of pitch in modulating emotional perception. Timbre, which refers to the tonal quality of sound, provides additional differentiation among emotions. Brighter timbres, characterized by complex harmonic structures, are linked to energetic emotions like excitement and happiness. Darker timbres, with simpler harmonic profiles, are associated with calmer or melancholic emotions like sadness [37] [44]. For instance, sadness often features very low timbre values, whereas happiness and anger are characterized by brighter timbres. Rhythm is another critical factor, as faster tempos are correlated with high-energy states such as happiness and excitement, while slower tempos correspond to low-energy emotions like sadness and calmness. This relationship is consistent across studies and reflects the temporal structure's influence on perceived arousal [37] [45].

By integrating these feature-emotion mappings, the Emotion Table serves as a foundation for systematically classifying and weighting emotional attributes in music. This approach ensures that each emotion is represented through a distinct profile of musical features, facilitating nuanced emotional classification.

#### Weighted Emotion Classification

*Feature Alignment*

Each song’s linguistic feature values, derived from audio feature extraction, were compared against the predefined Emotion Table, which maps emotions (e.g., Happy, Sad, Angry) to specific linguistic feature profiles. The alignment process determined the degree of similarity between the song’s features and the target feature ranges for each emotion. The strength of alignment was calculated based on the proximity of the song’s features to the target ranges specified in the Emotion Table.

*Salience Weighting*

To emphasize the significance of dominant features in emotional classification, a salience weighting mechanism was applied. Features that had stronger alignment with an emotion’s target range were assigned higher weights, reflecting their greater contribution to the emotion.

*Normalization*

The calculated weights for all emotions were normalized for each song to ensure probabilistic consistency. The sum of all weights for a single song was constrained to equal one, allowing the weighted labels to represent probabilities. This normalization step ensured that each song’s emotional profile was both interpretable and mathematically consistent.

*Output Labels*

The final emotional profile for each song consisted of weighted probabilities across multiple emotions, reflecting the degree to which each emotion was present in the song. These labels provided a multidimensional representation of the song’s emotional content, enabling more precise and flexible downstream applications.

## 3.4 Chatbot Emotion Prediction

The methodology used for emotion detection in chatbot interactions involved fine-tuning a prebuilt transformer-based model to classify user inputs into predefined emotional categories. The pretrained base model, "j-hartmann/emotion-english-distilroberta-base", was adapted to improve performance on the specific task using a labelled dataset. This process included data preparation, tokenization, model fine-tuning, and evaluation. By leveraging a pretrained language model and applying transfer learning, the system achieves robust emotion detection of predefined emotion categories.

#### Pretrained Model Overview

The chatbot’s emotion detection relies on the "j-hartmann/emotion-english-distilroberta-base" model, a transformer-based architecture pretrained for emotion classification. Built on DistilRoBERTa, a distilled version of the robust RoBERTa model, it offers computational efficiency and high accuracy in text-based tasks. The model leverages transformer layers to capture contextual relationships within text, making it well-suited for detecting nuanced emotional states. The model was pretrained on a diverse and balanced subset of text emotion datasets, with nearly 20,000 samples distributed across seven emotions: anger, disgust, fear, joy, neutral, sadness, and surprise. These datasets include text samples from various sources, such as Twitter posts, Reddit comments, self-reports, and TV dialogue utterances, ensuring wide-ranging emotional contexts [46].

#### Model Initialization

The pretrained DistilRoBERTa model was selected due to its proven capability in emotion detection tasks. The associated tokenizer was initialized to preprocess textual inputs into model-compatible formats by converting them into token IDs and attention masks.

#### Dataset Splitting

The consolidated and preprocessed dataset was split into two subsets to support the training and evaluation processes:

* Training Set (80%): Used to train the model and update its parameters.
* Validation Set (20%): Used to evaluate the model’s performance at the end of each epoch and ensure that the model generalizes well to unseen data.

The splitting ensured that no data leakage occurred between the training and validation sets.

#### Tokenization

Each text sample from the training and validation datasets was tokenized using the DistilRoBERTa tokenizer. The tokenization process included the following steps:

* Text Conversion: Text samples were tokenized into a sequence of input IDs.
* Attention Masks: Attention masks were generated to indicate which tokens should be attended to during training.
* Sequence Length Adjustment: All sequences were truncated or padded to a predefined maximum length to ensure uniformity in the input dimensions.

#### Fine-Tuning the Model

The fine-tuning process adapted the pretrained model to classify user inputs into the defined emotion categories (happy, excited, energetic, fear, sad, depressed, calm, angry). The following steps were involved:

*Supervised Training*

The model was trained in a supervised manner, where text samples (inputs) and their corresponding emotion labels (targets) were used to optimize the model’s weights.

*Loss Function*

A cross-entropy loss function was used to compute the error between the model’s predicted probabilities and the actual labels for each sample.

*Optimization*

The AdamW optimizer was employed to update the model’s weights during training. This optimizer was chosen for its ability to handle sparse gradients and improve convergence.

*Learning Rate Scheduling*

A learning rate scheduler was implemented to dynamically adjust the learning rate over the course of training, starting with a warm-up phase to stabilize training and gradually decreasing it to fine-tune the weights.

*Batch Processing*

The dataset was processed in batches to efficiently utilize computational resources. Batch size was chosen based on the hardware constraints and the model's memory requirements.

*Epoch-Based Training*

Training was conducted over multiple epochs, where the entire training dataset was passed through the model once per epoch. At the end of each epoch, the model’s performance was evaluated on the validation set.

#### Validation and Evaluation

During the fine-tuning process, the validation set was used to evaluate the model's performance at the end of each epoch. Accuracy was calculated as the primary metric to monitor how well the model generalized to unseen data. Once the training process was complete, the final weights of the fine-tuned model were saved to ensure that the best-performing version was preserved for deployment.

## 3.5 Chatbot Implementation

The chatbot was developed to facilitate emotionally aware, natural language conversations by integrating advanced NLP frameworks and emotion detection techniques. The implementation aimed to combine modular dialogue management with emotion analysis to create a robust system capable of providing personalized and contextually relevant responses. This section outlines the systematic process of building the chatbot, including the tools, methodologies, and justifications for design choices, highlighting how each component contributed to achieving the project’s objectives.

Apologies for my mistake. You clearly stated that you fine-tuned the emotion detection model. I’ll rewrite the methodology, fully reflecting your process and emphasizing that the fine-tuned model is your own work. Thank you for pointing this out; I will correct it immediately.

Methodology: Chatbot Implementation

Objective

The chatbot was developed to facilitate emotion-aware, dynamic, and contextually relevant conversations. It integrates a fine-tuned emotion detection model with natural language processing (NLP) tools to provide personalized user interactions. This methodology outlines the implementation process, focusing on the integration of the emotion detection model, dialogue management, and response generation.

#### Framework and Tools

*Rasa Framework:*

The Rasa framework was selected as the core conversational platform for managing intents, entities, and dialogue flows. Its modular architecture allowed seamless customization and integration of the fine-tuned emotion detection model with advanced policies for dynamic conversation management.

* **Rasa NLU**: Responsible for intent classification and entity extraction. Rasa NLU processes user inputs and categorizes them into intents, enabling the chatbot to understand the user's intention.
* **Rasa Core:** Manages dialogue flow and decision-making using machine learning policies and predefined rules. It predicts the next best action in a conversation based on context and the conversation’s history.

*Fine-Tuned Emotion Detection Model:*

The chatbot leverages a fine-tuned version of the "j-hartmann/emotion-english-distilroberta-base" model, specifically trained on curated emotion-labeled datasets to classify text into predefined categories (happy, excited, energetic, fear, sad, depressed, calm, angry). The fine-tuning process adapted the pretrained model to meet the chatbot's specific emotion detection requirements.

*OpenAI GPT-3.5 API:*

OpenAI’s GPT-3.5 API was used for generating natural, contextually aware responses. This ensured fluid and engaging conversations that adapt to user emotions.

#### Implementation Process

1. Intent and Entity Recognition

The chatbot was designed to identify user intents and extract entities from input text.

* NLU Pipeline: The Rasa NLU pipeline included the DIETClassifier for intent recognition and entity extraction, combined with SpacyNLP for tokenization and feature extraction. This pipeline ensured accurate mapping of user inputs to intents such as greet, goodbye, emotion\_query, and user\_message.
* NLU Training Data: The NLU training data consisted of diverse examples of user input mapped to intents such as greet, goodbye, emotion\_query, and user\_message.

2. Dialogue Management

Dialogue management was achieved using Rasa’s dialogue policies and flow customization capabilities.

* Rules and Stories: Rules defined fixed conversational paths for intents like greet and fallback, ensuring structured responses for simple interactions. Stories modeled dynamic and contextual dialogues, allowing multi-turn conversations to adapt based on user inputs.
* Policies: A combination of dialogue policies was employed to predict the next action based on conversation history:
  + RulePolicy for deterministic paths.
  + MemoizationPolicy for recalling previously encountered conversation patterns.
  + TEDPolicy for handling dynamic, context-aware responses using machine learning.

3. Emotion Detection Integration

The fine-tuned emotion detection model was integrated into the chatbot to analyze the emotional tone of user inputs:

* Preprocessing: User inputs were tokenized using the DistilRoBERTa tokenizer, generating input IDs and attention masks compatible with the model.
* Emotion Classification: The fine-tuned model predicted weighted probabilities for each of the eight emotion categories (happy, excited, energetic, fear, sad, depressed, calm, angry). These probabilities were normalized to ensure the total summed to one, creating a weighted emotional profile for each user message.

4. Conversational Response Generation

Dynamic and contextually appropriate responses were generated based on user inputs and detected emotions:

* Context Summarization: User inputs from the last five conversational turns were summarized to maintain context, particularly in multi-turn dialogues. This ensured the chatbot’s responses remained coherent and relevant to the ongoing conversation.
* Response Generation via GPT-3.5: The summarized context and detected emotional profile were sent to the OpenAI GPT-3.5 API, which generated responses tailored to the user's emotional state and conversational flow.

5. Integration with External Systems

The chatbot extended its utility beyond conversations by integrating with external systems:

* Emotion data generated by the chatbot was transmitted via a Flask API to external applications, such as a music recommendation engine.

## 3.6 Website Development

The website was developed as the primary interface for users to interact with the emotion-aware chatbot, analyze emotions, and access personalized music recommendations. Its design focused on integrating real-time chatbot interactions, emotion visualization, and seamless playlist management functionalities, ensuring a responsive and user-friendly experience.

#### Framework and Tools

*Frontend Framework*

React.js: Chosen for its modularity, component reusability, and efficient rendering using a virtual DOM. React ensured the development of a dynamic, responsive user interface (UI) optimized for performance. React.js was selected for its modular architecture, enabling efficient development and maintenance of complex, interactive components. Its ability to handle dynamic updates ensured smooth transitions between chatbot interactions and emotion visualization.

Bootstrap: Incorporated for consistent styling and responsive layouts across devices, expediting the UI/UX design process.

*Backend Framework*

Flask: Selected for its lightweight yet powerful architecture to handle server-side functionalities, API routing, and integration with external services like the chatbot and Spotify API. Flask’s lightweight architecture was ideal for handling multiple API calls, facilitating communication between the frontend, chatbot, and Spotify API. Its scalability supports future enhancements without significant architectural changes.

*APIs and Services*

Rasa REST API: Facilitated real-time communication between the chatbot and the website by processing user inputs and generating chatbot responses.

Spotify API: Enabled playlist creation, track management, and retrieval of music recommendations based on emotion analysis.

#### Implementation Process

1. Frontend Development

The frontend was designed with a focus on user interactivity, modularity, and real-time updates.

* Page Components:
  + Home Page: Introduced the VibeSync platform with an overview of features, engaging visuals, and a call-to-action for chatbot interaction.
  + Chatbot Page: Included the chat interface, emotion analysis visualization, and interactive features to ensure seamless user engagement.
  + Explore Page: Displayed emotion-classified songs using the Valence-Arousal model quadrants (e.g., high energy-positive or low energy-negative).
  + Playlist Management Page: Allowed users to input Spotify playlists, process them for emotion-based analysis, and save custom playlists.
* Interactive Features:

The chatbot interface dynamically updated with user and chatbot messages in real time. Emotion detection results were displayed using intuitive bar charts, highlighting the intensity of each detected emotion. Spotify song recommendations were embedded directly within the UI for immediate user access.

2. Backend Development

The backend was implemented using Flask to handle API requests, manage communication between components, and facilitate playlist management.

* Chatbot Integration: User messages were forwarded to the Rasa chatbot via Flask routes, ensuring real-time interaction. Chatbot responses, along with emotion profiles, were sent back to the frontend for display.
* Emotion Data Management: Detected emotion profiles were normalized and stored in a structured format. Flask endpoints allowed retrieval of these profiles for visualization on the website.
* Playlist Management System: Spotify playlists were analyzed for emotion alignment through Flask APIs. Routes for adding, retrieving, and saving playlists were implemented using Spotify’s Web API. Functions like to extract track features, map them to emotion categories, and recommend personalized playlists.

3. Integration of Emotion-Based Recommendations

Emotion-based recommendations were a core functionality of the website.

* Emotion Quadrant Mapping: Songs were categorized into four quadrants (e.g., high arousal-positive valence) using the Valence-Arousal model. Flask APIs facilitated retrieval of emotion-mapped songs. Recommendations were dynamically displayed on the Explore Page, allowing users to browse songs aligned with their detected emotions.
* Weighted Emotion Profiles: Emotion profiles from the chatbot were used to calculate song alignment scores. Flask handled cosine similarity calculations for matching songs to user emotions.
* Playlist Customization: Users could save curated playlists directly to their Spotify accounts through API integrations managed by Flask.

4. Key Features of the Website

* Chat Interface:
  + Integrated chatbot functionalities, enabling real-time, emotion-aware interactions.
  + Displayed detected emotions alongside conversation history for transparency.
* Emotion Visualization:
  + Generated bar charts to present detected emotions and their intensities.
  + Updated dynamically with each user input.
* Explore Page:
  + Displayed emotion-classified songs in quadrants, providing users with a structured view of recommended tracks.
  + Enabled navigation between quadrants to explore music for different emotional states.
* Playlist Management:
  + Allowed users to analyze existing Spotify playlists and create new ones based on emotion-based recommendations.
  + Provided summaries of playlist features, including average valence, arousal, and dominant emotional tones.

# 4.0 Results & Discussion

The Results section presents the outcomes of implementing the VibeSync chatbot and its associated website. The findings are structured to align with the objectives outlined in the methodology, showcasing the chatbot's performance, emotion detection accuracy, website functionality, and the effectiveness of emotion-based music recommendations. This section includes quantitative metrics, qualitative observations, and visual evidence to validate the methodologies used.

## 4.1 Valence-Arousal Prediction

This subsection presents the results of the valence and arousal prediction models. These models map Spotify features to emotional dimensions based on the Valence-Arousal (V-A) model, validating the methodology through performance metrics and model evaluations.

### 4.1.1 Valence Prediction Results

Linear Regression was selected as the model for valence prediction, primarily due to its simplicity, interpretability, and reasonable performance. The model achieved an R² score of 0.1460, an RMSE of 2.3209, and a MAE score of 2.1144. While Gradient Boosting demonstrated a higher R² score, Linear Regression was ultimately chosen for its computational efficiency and its ability to facilitate a straightforward implementation of the scaling process, making it a practical and effective choice for remapping Spotify valence values to the negative-positive range required by the Russell model.

The Linear Regression model produced the following scaling equation to transform Spotify valence values (ranging from 0 to 1) into the negative-positive range required by the Russell model:

#### Model Fit Analysis

The relationship between Spotify valence and scaled valence values is depicted in Figure 1. The scatterplot of training data points demonstrates a positive linear trend, with the fitted regression line clearly capturing the general direction of the data. The shaded confidence interval reflects the variability of the predictions, showing broader intervals at extreme valence values. This visualization confirms the effectiveness of the scaling equation in transforming Spotify valence into the desired negative-positive range.

A graph showing a number of dots

Description automatically generated with medium confidence

Figure 8: Linear Regression model fit for valence scaling, showing the relationship between Spotify valence and scaled valence, with a 95% confidence interval.

#### Residual Analysis

A collage of graphs

Description automatically generated

Figure 9: Residual diagnostics for the Linear Regression model, including Residuals vs. Fitted, Normal Q-Q Plot, Residual Distribution, and Scale-Location Plot.

Residual diagnostics for the Linear Regression model are shown:

* Residuals vs. Fitted Plot: The residuals are randomly scattered around zero, indicating that the model assumptions of linearity and no systematic error are satisfied.
* Normal Q-Q Plot: The points align closely with the diagonal line, suggesting that the residuals are approximately normally distributed, which is a key assumption for the Linear Regression model.
* Residual Distribution: The histogram of residuals reveals a near-normal distribution, further supporting the validity of the normality assumption.
* Scale-Location Plot: The plot shows a relatively even spread of residuals across the predicted values, indicating homoscedasticity (constant variance).

These residual plots collectively validate the Linear Regression model's suitability for the valence prediction task, confirming that no significant violations of the underlying assumptions occurred during model training and testing.

### 4.1.2 Arousal Prediction Results

The XGBoost model was selected for arousal prediction due to its superior performance in capturing the complex relationships between Spotify features and arousal levels. The model achieved an R² score of 0.8163, an RMSE of 1.0469, and an MAE of 0.6722, highlighting its robustness and reliability for this task.

#### Model Fit Analysis

A graph with a line

Description automatically generated

Figure 10: Predicted vs. actual arousal values for the training dataset, demonstrating the close alignment of predictions with the actual values.

The relationship between the predicted and actual arousal values for the training dataset is visualized in the figure. The scatterplot shows that the predicted arousal values closely follow the perfect-fit line (dashed red line), indicating a high degree of alignment between the model's predictions and the actual values. This alignment demonstrates the model's effectiveness in generalizing the arousal prediction task.

#### Residual Analysis

A graph of a number of blue dots

Description automatically generated with medium confidence

Figure 11: Residual diagnostics for the XGBoost model, including Residuals vs. Fitted and Residual Distribution.

Residual diagnostics for the XGBoost model are displayed as shown:

* Residuals vs. Fitted Plot: The residuals are randomly scattered around zero, indicating that the model effectively captures the underlying patterns without systematic errors.
* Residual Distribution: The histogram of residuals reveals a near-normal distribution, further supporting the validity of the model assumptions.

These plots confirm that the XGBoost model performs consistently well, with minimal bias or variance in its predictions.

#### Feature Importance

A screenshot of a computer

Description automatically generated

Figure 12: Relative importance of Spotify features in the arousal prediction task, highlighting the dominant role of Sp\_Energy.

The contribution of individual Spotify features to arousal prediction is presented. Among the features, Sp\_Energy was the most significant predictor, followed by Sp\_Speechiness and Sp\_Danceability, reflecting the importance of energy-related attributes in arousal modeling. Studies in music emotion recognition have consistently identified energy as a primary factor influencing arousal, with high-energy features correlating with heightened emotional intensity and excitement in listeners [33]. This insight aligns with the theoretical understanding that energy and intensity are strongly associated with arousal levels in music.

#### Testing Performance

A graph with a red line and blue dots

Description automatically generated

Figure 13: Predicted vs. actual arousal values for the test dataset, showcasing the model's generalization capabilities.

The performance of the model on the test dataset is illustrated above. The scatterplot shows that the predicted arousal values for the test data closely align with the actual values, maintaining consistency in the model's performance on unseen data. The slight deviations are within acceptable limits, further validating the robustness of the XGBoost model.

## 4.2 Weighted Emotion Classification

The results for weighted emotion classification were analyzed based on the classification of songs into emotional quadrants derived from the Valence-Arousal model. The classification process aimed to align each song's weighted emotional profile with one or more of the predefined emotion quadrants:

* Quadrant 1: High Arousal + Positive Valence (Happy, Excited, Energetic)
* Quadrant 2: High Arousal + Negative Valence (Fear, Angry)
* Quadrant 3: Low Arousal + Negative Valence (Sad, Depressed)
* Quadrant 4: Low Arousal + Positive Valence (Calm)

#### Model Performance

To evaluate the classification system, the model’s predictions for each song’s emotion quadrant were compared against the actual quadrant labels. The process used the average scores of emotional categories assigned to each quadrant to classify songs based on their highest-scoring quadrants. Songs could also be assigned to up to two quadrants based on their top scores. The final accuracy of the model, measured as the percentage of songs where the actual quadrant matched one of the predicted quadrants, was 64.81%. This indicates that the classification system correctly matched the true emotional quadrants for nearly two-thirds of the songs.

#### Methodology for Classification

* Quadrant Scoring:

Each song's weighted emotion scores were grouped based on the corresponding quadrant weights. For example, Quadrant 1 was computed as the average of the weighted scores for Happy, Excited, and Energetic emotions.

* Prediction of Quadrants:

The top one or two quadrants with the highest scores were selected as the predicted quadrants for each song. This allowed flexibility in cases where songs exhibited overlapping emotional characteristics.

* Accuracy Calculation:

A binary metric was used to determine whether the actual quadrant matched any of the predicted quadrants for each song. The Percentage of True Matches was computed as the proportion of songs where the actual quadrant was present in the predicted quadrants.

#### Limitations

The primary limitation of the weighted emotion classification system lies in its reliance on averaged scores for quadrant assignment, which may oversimplify complex emotional overlaps. Songs exhibiting emotional characteristics that fall near the boundaries of quadrants, such as those between Quadrant 1 (Happy, Excited, Energetic) and Quadrant 4 (Calm), are prone to misclassification due to the averaging process. Additionally, the model's assumption of equal importance for emotions within a quadrant (e.g., treating Happy and Energetic as equally weighted for Quadrant 1) may not accurately reflect the nuanced contributions of individual emotions. The classification system’s accuracy of 64.81% highlights these challenges, particularly in cases where emotional variability within a song is not fully captured by the weights or where songs exhibit characteristics of multiple quadrants beyond the top two predicted. Furthermore, the binary evaluation metric, whether the actual quadrant matched the top two predicted quadrants, does not account for partial correctness, potentially underestimating the model's capability to capture subtle emotional overlaps. These limitations suggest a need for refined weighting mechanisms, more sophisticated emotion modelling approaches, or additional features to improve the robustness and accuracy of the classification.

## 4.3 Chatbot Emotion Prediction

The performance of the fine-tuned emotion detection model integrated into the chatbot was evaluated to assess its effectiveness in identifying emotions from user inputs. This section presents the results of the model's evaluation, including detailed classification metrics, analysis of the confusion matrix, and an exploration of the model's strengths and limitations.

#### Model Evaluation Metrics

The fine-tuned model achieved an overall accuracy of **89.7%**, with an average prediction confidence of **0.908**. These results indicate the model's reliability in accurately predicting emotions across diverse text inputs. The detailed classification metrics, including precision, recall, and F1-score for each emotion category, provide further insights into the model's performance. Notably, the model demonstrated strong performance in detecting **Energetic** (F1-score: 0.950), **Depressed** (F1-score: 0.994), and **Calm** (F1-score: 0.946), reflecting its ability to identify both high-arousal and low-arousal emotions effectively. The macro-average precision, recall, and F1-score were all **0.897**, signifying balanced performance across all eight emotion categories. This balance indicates that the model does not disproportionately favor certain emotions over others, ensuring fairness and consistency in its predictions. The combination of high accuracy, confidence, and balanced macro-average metrics underscores the robustness of the emotion detection model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Emotion** | **Precision** | **Recall** | **F1-Score** | **Support** |
| **Happy** | 0.864 | 0.851 | 0.857 | 15,000 |
| **Excited** | 0.910 | 0.918 | 0.914 | 15,000 |
| **Energetic** | 0.937 | 0.963 | 0.950 | 15,000 |
| **Fear** | 0.784 | 0.851 | 0.816 | 15,000 |
| **Sad** | 0.874 | 0.790 | 0.830 | 15,000 |
| **Depressed** | 0.997 | 0.992 | 0.994 | 15,000 |
| **Calm** | 0.969 | 0.925 | 0.946 | 15,000 |
| **Angry** | 0.850 | 0.883 | 0.866 | 15,000 |

Table 8: Classification metrics for each emotion category, including precision, recall, F1-score, and support.

This table summarizes the **classification performance metrics** for each emotion category evaluated by the chatbot’s emotion detection model. The metrics include **Precision**, **Recall**, **F1-Score**, and **Support**, which provide a comprehensive understanding of the model's predictive capabilities for each emotion.

* **Precision**: Indicates the proportion of correctly predicted instances for an emotion out of all predictions made for that emotion.
* **Recall**: Reflects the proportion of correctly identified instances of an emotion out of all actual instances of that emotion.
* **F1-Score**: Represents the harmonic mean of precision and recall, offering a balanced measure of the model’s performance for each emotion. High F1-scores)indicate that the model performs well in both precision and recall for these emotions.
* **Support**: Refers to the number of samples in the dataset for each emotion. In this evaluation, all emotions were represented equally with 15,000 samples, ensuring balanced metrics across categories.

#### Confusion Matrix Analysis

A graph with blue squares

Description automatically generated

Figure 14: Confusion matrix illustrating true vs. predicted emotion categories, highlighting areas of high accuracy and emotional overlap.

The confusion matrix provides a detailed overview of the chatbot's emotion detection model performance by comparing the predicted labels against the true emotion labels for each category. The matrix not only highlights areas where the model excels but also points out potential overlaps and misclassifications between similar emotional states.

* High Accuracy for Certain Categories:
  + Depressed: The model demonstrated exceptional performance for the Depressed category, correctly predicting 14,874 out of 15,000 samples. This suggests the model can reliably identify this emotion, likely due to its distinct linguistic patterns.
  + Energetic and Excited: These categories also showed strong performance, with correct predictions for 14,439 and 13,768 samples, respectively, indicating the model's robustness in recognizing high-arousal positive emotions.
* Emotional Overlaps:
  + Fear and Angry: There is noticeable confusion between Fear and Angry, with 532 samples of Fear misclassified as Angry. This overlap likely stems from shared linguistic expressions of intensity and negativity, such as phrases indicative of heightened emotional states.
  + Sad and Fear: Some overlap was observed between Sad and Fear, with 711 instances of Fear misclassified as Sad. These categories share low-valence, emotionally negative tones, which can cause the model to misinterpret subtle textual differences.
* Boundary Confusion:
  + Calm and Sad: The model sometimes confuses Calm and Sad emotions, with 79 samples of Calm being misclassified as Sad. This could be attributed to overlapping linguistic indicators, especially when emotional tones are subdued.
  + Happy and Excited: These categories show minor misclassifications (e.g., 662 samples of Happy classified as Excited) due to shared positive and high-arousal linguistic features.
* Rare Misclassifications:
  + Misclassifications between dissimilar emotions, such as Happy being predicted as Depressed or Energetic being predicted as Calm, are minimal. This indicates the model's ability to distinguish between starkly different emotional tones with high accuracy.

# 5.0 Conclusion

The VibeSync project aimed to bridge the gap between emotion-based music recommendation systems and dynamic user interactions by integrating advanced machine learning techniques, a fine-tuned emotion detection model, and a chatbot interface. The system demonstrated significant potential in addressing the limitations of existing music recommendation frameworks by leveraging weighted emotion classification and real-time emotion detection through conversational interactions.

The methodology employed in this study allowed for the accurate classification of songs and user emotions. The fine-tuned emotion detection model achieved robust performance metrics, enabling the chatbot to dynamically detect user emotions across eight categories. Similarly, the weighted emotion classification method provided nuanced song categorizations, ensuring that recommendations were not only aligned with the user’s emotional state but also versatile and scalable for future extensions. The inclusion of Valence-Arousal predictions further enriched the system's analytical foundation by mapping songs onto an emotional circumplex, enabling a more holistic approach to mood-based music recommendations.

While the system achieved its intended objectives, certain limitations persisted. These include challenges in achieving perfect emotion detection due to subjective ambiguities, computational trade-offs in selecting lightweight models for efficiency, and constraints in balancing real-time performance with predictive accuracy. Nevertheless, the results achieved—such as an emotion detection accuracy of 89.7% and a valence prediction model exhibiting reasonable RMSE—underline the system's capacity to revolutionize user-centric music experiences.

In conclusion, VibeSync represents a significant advancement in emotion-based music recommendation systems. Its modular design, combining an emotion-aware chatbot and scalable recommendation framework, positions it as a promising solution for enhancing personalized user experiences. Future work can explore the integration of additional modalities, such as facial expression or voice analysis, and further refine the algorithms for real-time responsiveness, ensuring even broader applicability and user satisfaction.

#### Challenges

Throughout the development of the VibeSync system, several challenges emerged that influenced the design, implementation, and functionality of the project. These challenges spanned technical, operational, and external dependency factors, impacting various aspects of the system.

* Spotify API Restrictions and Changes

A significant challenge arose due to Spotify removing certain key features, such as song recommendations and similar playlist functionality, from their public API. These features were crucial for enhancing the personalization and scalability of the music recommendation system. Their removal necessitated a shift in approach, requiring alternative methods to generate recommendations, such as leveraging pre-existing datasets and fine-tuning the recommendation logic within the system.

* Complexity in Feature Extraction

The reliance on Spotify's API to retrieve audio features, such as energy, valence, and tempo, presented additional challenges. Despite the richness of Spotify's audio feature set, its API limitations meant that bulk queries for large datasets were impractical due to rate limits and latency issues. This added overhead to the development process, requiring optimizations to manage API calls efficiently and implement caching mechanisms to reduce repeated requests.

* Data Availability and Processing

While multiple datasets were utilized for fine-tuning the emotion detection model, ensuring uniformity across these datasets in terms of emotion categories and linguistic contexts was challenging. Preprocessing steps, such as balancing the dataset and removing irrelevant categories, introduced complexities that required careful handling to avoid data biases. Additionally, extracting features for songs not included in Spotify's database required additional processing using tools like Librosa, adding computational and storage overheads.

# References

[1] R. I. M. Dunbar, K. Kaskatis, I. MacDonald, and V. Barra, “Performance of music Elevates pain threshold and positive affect: Implications for the Evolutionary Function of music,” Evolutionary Psychology, vol. 10, no. 4, pp. 688–702, Oct. 2012, doi: 10.1177/147470491201000403.

[2] A. Nawaz and A. Rana, “Impact of music on Mood: Empirical investigation,” SSRN Electronic Journal, Nov. 2015, [Online]. Available: https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=2696883

[3] Y. Yang and Y. Sun, “Lyrically Yours: A Mobile Application for Automated Music Therapy through Lyric Analysis Utilizing Natural Language Processing and Machine Learning,” Software Engineering & Trends, Apr. 2024, pp. 215–225, Apr. 2024, doi: 10.5121/csit.2024.140820.

[4] A. Nair, S. Pillai, G. S. Nair, and A. T, “Emotion Based Music Playlist Recommendation System using Interactive Chatbot,” 2022 7th International Conference on Communication and Electronics Systems (ICCES), pp. 1767–1772, Jul. 2021, doi: 10.1109/icces51350.2021.9489138.

[5] T. Tofalvy and J. Koltai, “‘Splendid Isolation’: The reproduction of music industry inequalities in Spotify’s recommendation system,” New Media & Society, vol. 25, no. 7, pp. 1580–1604, Jul. 2021, doi: 10.1177/14614448211022161.

[6] Dr. C. . M. Rao, “Hybrid Deep Learning Approach to Emotion-Infused Music recommendation,” INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, vol. 08, no. 04, pp. 1–5, Apr. 2024, doi: 10.55041/ijsrem30959.

[7] A. D. R. N, P. Karlapati, M. R. Mulagondla, P. Amaranayani, and A. P. Toram, “An innovative emotion recognition and solution recommendation chatbot,” 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 1100–1105, Mar. 2022, doi: 10.1109/icaccs54159.2022.9785269.

[8] J. Zhang, Y. J. Oh, P. Lange, Z. Yu, and Y. Fukuoka, “Artificial intelligence Chatbot Behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: Viewpoint,” Journal of Medical Internet Research, vol. 22, no. 9, p. e22845, Sep. 2020, doi: 10.2196/22845.

[9] K. S. Krupa, G. Ambara, K. Rai, and S. Choudhury, “Emotion aware Smart Music Recommender System using Two Level CNN,” 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1322–1327, Aug. 2020, doi: 10.1109/icssit48917.2020.9214164.

[10] M. Susino and E. Schubert, “Cultural stereotyping of emotional responses to music genre,” Psychology of Music, vol. 47, no. 3, pp. 342–357, Mar. 2018, doi: 10.1177/0305735618755886.

[11] M. R. Dalida, L. B. Aquino, W. C. Hod, R. A. Agapor, S. L. Huyo-A, and G. A. Sampedro, “Music mood prediction based on Spotify’s audio features using logistic regression,” 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Dec. 2022, doi: 10.1109/hnicem57413.2022.10109396.

[12] K. Machová, M. Szabóova, J. Paralič, and J. Mičko, “Detection of emotion by text analysis using machine learning,” Frontiers in Psychology, vol. 14, Sep. 2023, doi: 10.3389/fpsyg.2023.1190326.

[13] M. Karna, D. S. Juliet, and R. C. Joy, “Deep learning based Text Emotion Recognition for Chatbot applications,” 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 988–993, Jun. 2020, doi: 10.1109/icoei48184.2020.9142879.

[14] F. Zhuang et al., “A comprehensive survey on transfer learning,” arXiv (Cornell University), Jan. 2019, doi: 10.48550/arxiv.1911.02685.

[15] K. S. Krupa, G. Ambara, K. Rai, and S. Choudhury, “Emotion aware Smart Music Recommender System using Two Level CNN,” 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1322–1327, Aug. 2020, doi: 10.1109/icssit48917.2020.9214164.

[16] R. Ma, Y. J. Colón, and T. Luo, “Transfer Learning Study of Gas Adsorption in Metal–Organic Frameworks,” ACS Applied Materials & Interfaces, vol. 12, no. 30, pp. 34041–34048, Jul. 2020, doi: 10.1021/acsami.0c06858.

[17] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, “Transfer learning using computational intelligence: A survey,” Knowledge-Based Systems, vol. 80, pp. 14–23, Jan. 2015, doi: 10.1016/j.knosys.2015.01.010.

[18] S. Xue, J. Lu, G. Zhang, and L. Xiong, “Heterogeneous feature space based task selection machine for unsupervised transfer learning,” 2021 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), vol. 25, pp. 46–51, Nov. 2015, doi: 10.1109/iske.2015.29.

[19] M. Kaminskas and F. Ricci, “Contextual music information retrieval and recommendation: State of the art and challenges,” Computer Science Review, vol. 6, no. 2–3, pp. 89–119, May 2012, doi: 10.1016/j.cosrev.2012.04.002.

[20] R. Delbouys, R. Hennequin, F. Piccoli, J. Royo-Letelier, and M. Moussallam, “Music mood detection based on audio and lyrics with deep neural net,” arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1809.07276.

[21] Y.-S. Seo and J.-H. Huh, “Automatic Emotion-Based Music classification for supporting intelligent IoT applications,” Electronics, vol. 8, no. 2, p. 164, Feb. 2019, doi: 10.3390/electronics8020164.

[22] Q. D. P. Bayu, S. Suyanto, and A. Arifianto, “Hierarchical SVM-KNN to classify music emotion,” 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 5–10, Dec. 2019, doi: 10.1109/isriti48646.2019.9034651.

[23] N. L. Lu, D. Liu, and N. H.-J. Zhang, “Automatic mood detection and tracking of music audio signals,” IEEE Transactions on Audio Speech and Language Processing, vol. 14, no. 1, pp. 5–18, Dec. 2005, doi: 10.1109/tsa.2005.860344.

[24] M. H. Nofal, Z.K.A Baizal, and R. Dharayani, “Multi Criteria Recommender System for Music using K-Nearest Neighbors and Weighted Product Method,” Ind. Journal on Computing, vol. 6, no. 2, Sep. 2021, doi: 10.34818/indojc.2021.6.2.575.

[25] F. H. Rachman, R. Sarno, and C. Fatichah, “Music Emotion Classification based on Lyrics-Audio using Corpus based Emotion,” International Journal of Electrical and Computer Engineering (IJECE), vol. 8, no. 3, p. 1720, Jun. 2018, doi: 10.11591/ijece.v8i3.pp1720-1730.

[26] S.-H. Chang, A. Abdul, J. Chen, and H.-Y. Liao, “A personalized music recommendation system using convolutional neural networks approach,” 2018 IEEE International Conference on Applied System Invention (ICASI), pp. 47–49, Apr. 2018, doi: 10.1109/icasi.2018.8394293.

[27] N. Mathew, N. Chooramun, and M. S. Sharif, “Implementing a Chatbot Music Recommender System based on User Emotion,” 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), pp. 195–199, Nov. 2023, doi: 10.1109/3ict60104.2023.10391771.

[28] R. De Prisco, A. Guarino, D. Malandrino, and R. Zaccagnino, “Induced Emotion-Based Music Recommendation through Reinforcement Learning,” Applied Sciences, vol. 12, no. 21, p. 11209, Nov. 2022, doi: 10.3390/app122111209.

[29] A. P. Rao, N. Nithasha, P. R, P. S, and R. Rao, “Emotune: Emotion and gender Aware Music Generation Chatbot,” IJARCCE, vol. 13, no. 4, Apr. 2024, doi: 10.17148/ijarcce.2024.134152.

[30] Rao, Nitasha, Prathik, and Rao, “EMOTUNE,” International Journal of Creative Research Thoughts (IJCRT), vol. 12, Jan. 2024, [Online]. Available: http://www.ijcrt.org/papers/IJCRT2401181.pdf

[31] P. Sarda, S. Halasawade, A. Padmawar, and J. Aghav, “Emousic: Emotion and Activity-Based Music Player using Machine Learning,” in Advances in intelligent systems and computing, 2019, pp. 179–188. doi: 10.1007/978-981-13-6861-5\_16.

[32] S. Gilda, H. Zafar, C. Soni, and K. Waghurdekar, “Smart music player integrating facial emotion recognition and music mood recommendation,” 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 154–158, Mar. 2017, doi: 10.1109/wispnet.2017.8299738.

[33] K. J. Lee, G.-E. Lee, S. H. Lee, and J.-H. Lee, “Differences in arousal and valence on the Korean phoneme of artificial voice between Korean and Chinese women,” PLoS ONE, vol. 18, no. 4, p. e0284045, Apr. 2023, doi: 10.1371/journal.pone.0284045.

[34] Malheiro R., Panda R., Gomes P., and Paiva R. P., “Bi-Modal Music Emotion Recognition: Novel Lyrical Features and Datase,” 9th International Workshop on Music and Machine Learning – MML’2016 – in Conjunction With the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2016, [Online]. Available: https://mir.dei.uc.pt/downloads.html#MERGE

[35] P. L. Louro, H. Redinho, R. Santos, R. Malheiro, R. Panda, and R. P. Paiva, “MERGE -- a bimodal dataset for static music emotion recognition,” arXiv (Cornell University), Jul. 2024, doi: 10.48550/arxiv.2407.06060.

[36] Z. Liu, A. Xu, Y. Guo, J. U. Mahmud, H. Liu, and R. Akkiraju, “Seemo,” Conference: The 2018 CHI Conference, pp. 1–12, Apr. 2018, doi: 10.1145/3173574.3173938.

[37] A. S. Bhat, V. S. Amith, N. S. Prasad, and D. M. Mohan, “An Efficient Classification Algorithm for Music Mood Detection in Western and Hindi Music Using Audio Feature Extraction,” 2014 Fifth International Conference on Signal and Image Processing, Bangalore, India, Jan. 2014, doi: 10.1109/icsip.2014.63.[38] “Web API Reference | Spotify for Developers.” https://developer.spotify.com/documentation/web-api/reference/get-audio-features

[39] Z. Liu, A. Xu, Y. Guo, J. U. Mahmud, H. Liu, and R. Akkiraju, “Seemo,” *2018 CHI Conference*, pp. 1–12, Apr. 2018, doi: 10.1145/3173574.3173938.

[40] M. Nuzzolo, “Music Mood Classification – Electrical and Computer Engineering Design Handbook,” Mar. 25, 2015. https://sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/

[41] X. Liu and Y. Xu, “Relations between affective music and speech: evidence from dynamics of affective piano performance and speech production,” *Frontiers in Psychology*, vol. 6, Jul. 2015, doi: 10.3389/fpsyg.2015.00886.

[42] D. Huron, “Affect induction through musical sounds: an ethological perspective,” *Philosophical Transactions of the Royal Society B Biological Sciences*, vol. 370, no. 1664, p. 20140098, Feb. 2015, doi: 10.1098/rstb.2014.0098.

[43] C. Nussbaum, A. Schirmer, and S. R. Schweinberger, “Musicality – Tuned to the melody of vocal emotions,” *British Journal of Psychology*, vol. 115, no. 2, pp. 206–225, Oct. 2023, doi: 10.1111/bjop.12684.

[44] A. M. Grimaud and T. Eerola, “An interactive approach to emotional expression through musical cues,” *Music & Science*, vol. 5, Jan. 2022, doi: 10.1177/20592043211061745.

[45] H. E. Kragness and L. J. Trainor, “Nonmusicians express emotions in musical productions using conventional cues,” *Music & Science*, vol. 2, p. 205920431983494, Jan. 2019, doi: 10.1177/2059204319834943.

[46] J. Hartmann, “"Emotion English DistilRoBERTa-base,” *Hugging Face*, 2022. https://huggingface.co/j-hartmann/emotion-english-distilroberta-base