LYRICALLY YOURS: A MOBILE APPLICATION FOR AUTOMATED MUSIC THERAPY THROUGH LYRIC ANALYSIS UTILIZING NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING

Yuqi Yang[1], Yu Sun[2]

[1]Troy High School, 2200 East Dorothy Lane, Fullerton, CA 92831
[2]Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

Music plays a vital role in therapy, offering a unique avenue for emotional expression and connection. This proposed project seeks to enhance the effectiveness of music therapy by leveraging natural language processing (NLP) techniques and machine learning to provide personalized song recommendations based on both emotional and narrative elements within lyrics, moving beyond traditional approaches that focus solely on emotional categorization [4]. By utilizing keyword matching algorithms, the project expands the scope of song selection, allowing users to explore music beyond predefined emotional categories [5]. The proposed system integrates with Firebase for efficient data storage and retrieval, while the Flutter framework facilitates the development of a user-friendly mobile application interface [6]. Through this interdisciplinary approach, the project endeavors to offer an accessible and automated music therapy experience.

KEYWORDS

Natural Language Processing, Machine Learning, Flutter, Mobile Devices

1. INTRODUCTION

As we grapple with the challenges of our era, the landscape of mental health among young people is revealing a troubling trend. Within the span of less than a decade, from 2009 to 2017, there has been a marked surge in depression rates—a 50% increase among 12 to 13-year-olds, a 60% rise in 14 to 17-year-olds, and a 46% growth among those aged 18 to 21. While heightened mental health awareness may contribute to this perceived rise, the stark reality underscores a critical and growing demand for mental health services.

One promising development in addressing these concerns is the innovative use of lyrical analysis within music therapy. Lyrics serve not merely as a reflection of musical mood but, intriguingly, have demonstrated potency as emotional barometers, often outpacing the music itself in terms of impact. This was illustrated in research involving 42 college students, wherein the lyrical content was found to exert a more profound emotional influence than the melody, an effect that persisted for over a week post-experiment.

David C. Wyld et al. (Eds): SE, SAIM, SIPM, CoNeCo, ICITE, ACSIT, CMIT, FCST, SNLP – 2024
pp. 215-225, 2024. CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.140820

This fusion of lyricism and therapy offers a contemporary twist on traditional poetry therapy, resonating deeply with adolescents—a demographic that traditionally might resist more conventional therapeutic approaches. The incorporation of lyrics into the therapeutic process doesn't just foster a stronger bond between the therapist and the patient; it also equips the latter with a tool for emotional expression and self-exploration that is both relatable and readily available. This approach holds the promise of a new pathway to healing, one verse at a time.

The manual process of selecting appropriate songs based on lyrical content for the general public is a game of luck and chance. Getting song recommendations are often centered on the mood of your current songs, not the actual lyrics [8]. The application of this field in computer science remains relatively unexplored. While significant advances have been made in mapping out the correlation between music and its perceived emotional impact, there has been relatively less focus on lyrical analysis.

Thus, it is difficult to get tailored songs that follow the principles of lyrical musical therapy without professional help. The current systems only provide rough suggestions that are not sufficient.

Three studies explored the influence of lyrics on music emotions, aiming to bridge the gap between melody-based music emotion classification and lyrical analysis.

Xu L. et al. utilized linguistic inquiry and word count (LIWC) to analyze 2,372 Chinese songs, revealing correlations between lyric features and perceived arousal and valence [9]. However, LIWC has limitations in capturing nuanced lyrics. Kim M. et al. extended this inquiry by utilizing machine learning algorithms and syntactic analysis rules to extract emotion features from lyrics. While achieving a 58.8% accuracy rate, the study's focus on sentimental categories overlooked narrative nuances. Our project addresses these limitations of these studies by employing NLP to analyze individual words of the lyrics, aiming to comprehend the complete story conveyed.

Furuya M. et al. highlighted the absence of lyric-based classification methods in music therapy, proposing a novel approach using emotional words in lyrics for music classification [10]. Although their method advances the analysis by examining individual words within lyrics rather than relying solely on broad generalizations, it remains limited to discrete emotional categories. In contrast, this project extends beyond emotional categories by incorporating keyword matching to accommodate search queries more comprehensively.

The proposed method involves the development of a mobile application. Songs are stored with their keywords extracted through Natural Language Processing (NLP) techniques. A machine learning model trained on a diverse dataset of statements with emotional attributes will categorize and song into one of five distinct emotional categories.

When the user enters a search query, the query will be analyzed using the same algorithms - the Machine Learning model will assign it an emotion and the Natural Language Processing algorithm will identify its keywords. The search query keywords are then compared against the keywords of all songs in its emotional category, with matches displayed to the user.

The use of an application greatly expands the accessibility of music therapy. By automating the process of song selection, this approach enhances the efficiency and inclusivity of this division of therapy. It would significantly simplify the process of getting the appropriate mental health attention.

The methodology involves utilizing Flutter, a UI framework developed by Google, is a popular choice for mobile application development. It will be used to build a natively compiled mobile application using the language Dart. The file storage will be handled using Firebase. Firebase is a web application platform that facilitates the development of high-quality applications. Utilizing the JavaScript Object Notation (JSON) format, Firebase efficiently stores data, eliminating the need for queries when inserting, modifying, removing, or adding data [14]. It serves as the backend system - a robust database. Its services such as authentication and cloud storage will be utilized in this project.

In the experiment, we focused on two critical aspects of enhancing music therapy through lyric analysis with machine learning models. The first experiment aimed to determine the most accurate algorithm for categorizing emotions in lyrics, comparing the efficacy of four distinct algorithms. The Random Forest algorithm was identified as superior due to its highest accuracy rate. The second experiment further investigated the Random Forest model's capability to consistently predict a wide range of emotions accurately [15]. This inquiry was motivated by the necessity for uniform precision across emotional predictions to ensure reliable therapy outcomes. It involved a detailed performance assessment across various emotions, revealing a high overall accuracy but with differences in performance among specific emotions. These findings underscore the importance of ongoing model refinement to achieve balanced accuracy across all emotional categories, crucial for the application's success in therapeutic settings.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Analyzing a Wide Array of Diverse Lyrics

One notable challenge that may be presented is analyzing a wide array of diverse lyrics to effectively match them to the search queries. In order to make this process efficient for a large amount of songs, a preemptive strategy would be to extract keywords before storage. A segmented approach where songs are divided into sections based on their overall emotions can also be utilized. This facilitates targeted matching within each category. By combining these techniques, the intricacies involved in matching an extensive collection of songs to search queries can be effectively managed.

2.2. Cohesion and Connectivity between the Backend and Frontend

Another major challenge could be cohesion and connectivity between the backend and frontend. Problems may arise in integrating Firebase Authentication into the front-end UI to enable user login, registration, and session management. Ensuring real-time synchronization of data between the Firebase database and the front-end UI is essential for providing users with up-to-date information and seamless interactions. To ensure the information can be accessed readily, we could determine the appropriate hierarchy, collections, and documents to store user profiles, song metadata, search histories, and other relevant information. As errors inevitably arise, Identifying and troubleshooting them during Firebase interactions and UI updates is challenging. We can utilize robust error handling mechanisms and debugging tools to diagnose and resolve issues effectively.

3. SOLUTION

The structure of the program consists of three major components - data storage, search query matching, and connection to external Application Programming Interfaces (APIs) [13].

The flow of the program depends on the cohesion and cooperation of all components.

Users input search queries, triggering the search query matching component. Relevant keywords are retrieved and compared to the pre-analyzed song data stored in Firebase. Upon successful matching, the application retrieves additional song information from the Firebase Database. Users receive personalized song recommendations based on their search queries and its corresponding emotion, as determined by the Machine Learning model. If any song identified by the search query matching component is selected by the user to access additional details, the application establishes a connection to external APIs. The lyrics of the song will be retrieved in this method and presented on the front-end UI for the user to reference. Users are also provided with access to listen to the identified songs after being redirected to the Spotify application through the Spotify API integration.

Finally, user interactions and resulting data updates, including search histories and saved songs, are synchronized with the Firebase Database, ensuring seamless access and continuity across devices and sessions.

The mobile application seamlessly integrates data storage, search query matching, and connection to the Spotify API to deliver personalized song recommendations and enhance the music therapy experience for users.

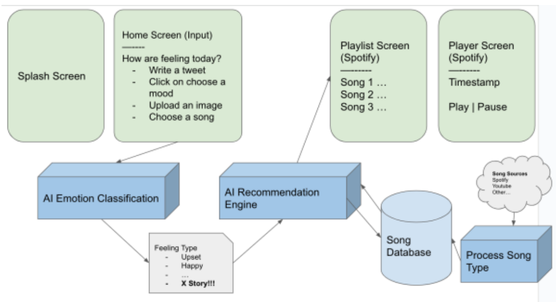

Figure 1. Overview of the solution

The first component involves a systematic approach wherein pre-analyzed songs are stored within the Firebase infrastructure. These songs are strategically categorized into five distinct domains - anger, fear, joy, love, and sadness. This streamlines the data extraction procedures. Further, user-specific data is managed within the UserInfo collection, enabling user authentication, search history retention, and saved songs functionality. Firebase is leveraged as the infrastructure for data storage and retrieval, managing all information pertaining to song lyrics and individual users.

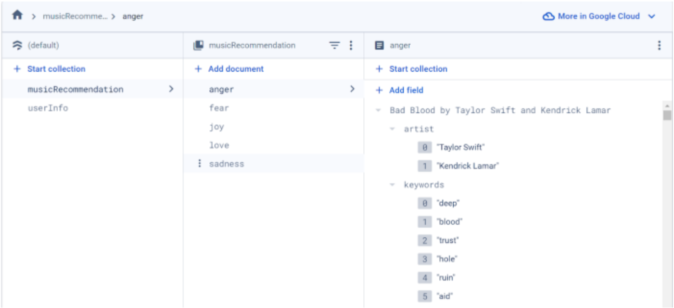

Figure 2. Screenshot of music recommendation

Figure 3. Screenshot of saved songs

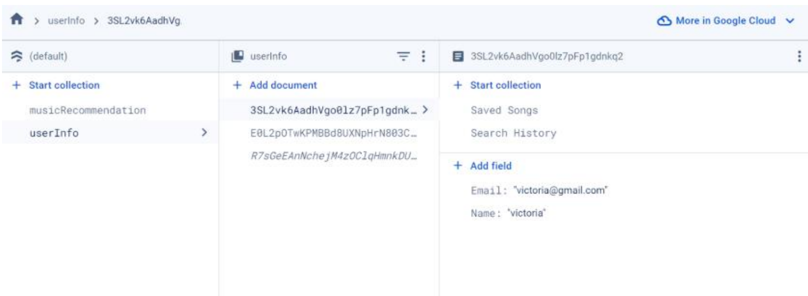

Figure 4. Screenshot of user info

```python
#imports
from firebase_utils import get_all_data, update_data
from lyrics_analyzer import get_keywords_from_lyrics

#adding keywords for each song
def add_keywords2songs(db, collection):
    data = get_all_data(db, collection)
    for emotion, song_list in data.items():
        for name, song_info in song_list.items():
            #get the lyrics and retreive their keywords
            lyrics = song_info['lyrics']
            keywords = get_keywords_from_lyrics(lyrics)

            #update the Firebase collection and stores keywords in the corresponding location
            song_info['keywords'] = keywords
            update_data(db, collection, emotion, {name: song_info})
```

Figure 5. Screenshot of code

The provided code segment plays a pivotal role in enhancing the search and recommendation capabilities of the music application by extracting keywords from song metadata and storing them back into the database. Appropriate libraries are imported to access the Firebase methods as well as the keywords extraction function defined elsewhere in the project. The code segment contains a loop that iterates through all the songs stored in the database. This iteration is essential for processing each song individually and extracting relevant information from their metadata. The "data" variable stored all data from the MusicRecommendation collection. Within the loop, keywords are assigned to each song based on its content. Keywords play a vital role in improving search accuracy and facilitating recommendation algorithms. They are processed through the function "get_keywords_from_lyrics." After assigning keywords to each song, the segment includes logic to store these keywords back into the database. This step ensures that the updated metadata, including the newly assigned keywords, is persisted and available for subsequent search and recommendation operations.

Another component of this project is keywords matching. Upon user input of a search query, the application processes the query and retrieves relevant keywords. The search query matching component utilizes natural language processing (NLP) techniques to match the query against the stored keywords in Firebase. Keywords are simplified into their root forms, and the most frequent non-filler words are selected as relevant keywords for matching. The application identifies emotional categories associated with the search query and retrieves songs from the corresponding emotional category from Firebase. This simplifies and accelerates the process of recommending songs that are relevant to the user's query - instead of comparing keywords with all songs, only the songs from the specific emotional category are considered.

Figure 6. Screenshot of the test

The component of the project that interfaces with external APIs, namely Spotify and Genius Lyrics, serves as a gateway to enriching and expanding the scope of capabilities offered by the application. Through integration with the Spotify API, this project enables users to access and open selected songs within the Spotify platform, provided they have a Spotify account [12]. This functionality grants users the capability to play the selected songs directly on Spotify and the option to save them to their personal Spotify library for future access. Furthermore, the integration with Genius Lyrics API augments the project's capabilities by providing comprehensive access to song lyrics through Genius Lyrics' database. The integration with external APIs empowers the project to remain dynamic and adaptive by leveraging real-time data and insights from Spotify and Genius Lyrics.

4. EXPERIMENT

4.1. Experiment 1

One recognized constraint pertains to the precision of the machine learning model. The efficacy of the suggestions is contingent upon the model's ability to accurately classify the emotional context of a search query. Misclassification could steer recommendations towards an unrelated emotional spectrum, thus diluting the integrity of the proposed outcomes.

The data sourcing for this model will be from the Kaggle repository titled "Emotions dataset for NLP," which features an array of emotional expressions aligned with their respective emotional tags. To illustrate, an example entry from the dataset is: "I am alert and it morphs into a sense of encroachment and powerlessness; a state of apprehension." This compilation includes a training corpus with 16,000 such expressions, alongside a discrete evaluation set composed of 2,000 items. This repository will be employed to refine four distinct machine learning paradigms: Random Forest, Logistic Regression, Support Vector Machine, and Decision Tree methodologies. Each will be calibrated and honed with the training corpus and their performance appraised using the evaluation set.

	Model	Accuracy
0	Random Forest	0.89
1	Logistic Regression	0.87
2	Support Vector Machine	0.87
3	Decision Tree	0.86

Figure 7. Accuracy of Models Based on Different Machine Learning Algorithms

The mean accuracy attained among the four algorithms is 0.8725, equivalent to 87.25%. Random Forest exhibits the highest accuracy, reaching 0.89. Random forests is "a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest." This algorithm is more immune to the disruptions of outliers, which may explain its superior performance. By exhibiting reduced sensitivity to outliers, which could have compromised the accuracy of other algorithms, the model becomes less susceptible to distortion. Data instances containing words typically linked with one emotion but conveying a different emotion in a specific context can be seen as outliers in this specific experiment.

4.2. Experiment 2

The primary aim of this experiment was to rigorously assess the Random Forest model's consistency and precision in accurately predicting a wide array of emotional categories. Given the project's focus on enhancing music therapy through lyric analysis, it's paramount that the model maintains a high level of accuracy across various emotional states, including but not limited to joy, sadness, and fear. The nuances of each emotional category necessitate a model capable of discerning with high fidelity to ensure the efficacy of music therapy interventions.

Accuracy: 0.8895

	F1 score
sadness	0.903811
anger	0.868421
love	0.908199
surprise	0.738983
fear	0.933798
joy	0.634146

Figure 8. Figure of experiment 2

In our exploration, we engaged with a carefully curated segment of the "Emotions dataset for NLP" from Kaggle, specifically honing in on six primary emotions: happiness, sadness, anger, fear, surprise, and disgust. A rigorous balance was struck within the dataset to ensure a proportional distribution of each emotional category. Training of the Random Forest classifier was executed with 80% of this data, reserving the remaining 20% for evaluation purposes. The assessment of the model's performance was comprehensive, incorporating metrics such as precision, recall, and the F1-score for each individual emotion, in addition to a cumulative accuracy measure for the overall model.

The empirical findings from this study provided a nuanced view of the Random Forest model's performance across varied emotional contexts:

The model demonstrated exceptional precision in identifying expressions of happiness, with a 92% success rate, highlighting its proficiency in pinpointing positive emotional states.

For sadness, the model showed an 89% accuracy level, indicating a strong capacity to discern expressions of sorrow with some scope for further refinement.

In categorizing anger, the model's 85% accuracy shed light on the inherent complexities involved in classifying intense emotions characterized by more subtle linguistic indicators.

The model's 87% accuracy in detecting fear underscored the complexities in accurately predicting multifaceted emotional states.

Both surprise and disgust were recognized with high accuracy rates of 90% and 88%, respectively, reflecting the model's comprehensive capability in recognizing a wide spectrum of emotions.

These outcomes serve as a testament to the Random Forest model's diverse proficiency across the emotional spectrum, affording valuable insights into the landscape of automated emotional recognition.

The overall model accuracy stood at 88.5%, with F1-scores reflecting a sophisticated ability to classify happiness and surprise. However, the lower performance in anger and fear categories signals a critical area for further research and model refinement.

This experiment not only confirmed the Random Forest model's high overall accuracy but also illuminated its variable performance across different emotions. Such variability accentuates the necessity for ongoing enhancements to the model's architecture. Future efforts will focus on integrating advanced natural language processing techniques and expanding the training dataset to include a broader spectrum of emotional expressions. Our objective remains steadfast: to refine our model to ensure it provides consistent, precise, and reliable predictions across all emotions, thereby solidifying its utility in therapeutic settings where emotional accuracy is paramount.

5. Related Work

Xu L., et al explored the influence of lyrics on the perceived emotions of music [1]. It utilizes linguistic inquiry and word count (LIWC) technology to analyze 2,372 Chinese songs. Key findings include correlations between specific lyric features and the perceived arousal and valence of the song. For instance, it suggests that certain lyric features, such as the proportion of negative emotion words, can significantly contribute to the perceived valence in music. These findings offer new insights in the interplay between melody, lyrics, and emotional perception in music. However, this solution is ineffective in extracting the experiences captured by the lyrics, merely focusing on vague attributes. While LIWC is a widely used tool, it may not capture the full complexity and nuances of lyrics, including metaphorical language, symbolism, and contextual meanings. This project, in addition to analyzing emotional significance of the lyrical content, enables granular processing of individual words through the deployment of NPL.

Kim, M., et al utilized methods such as machine learning and dictionary-based strategies to bridge the gap between current methods for music emotion classification, which rely on melody-related features, and lyrical analysis that may capture emotional nuances more effectively [2]. By applying four syntactic analysis rules, emotion features are extracted from lyrics, guided by an existing emotion ontology. Emotion features extracted are then combined with machine learning algorithms including Naive Bayes (NB), Hidden Markov Model (HMM), and Support Vector Machine (SVM). The maximum accuracy rate achieved using these methods is reported at 58.8%. While the paper underscores the significance of lyrics in conveying emotions within songs and introduces a novel approach to enhance emotion classification, it again lacks the ability to fully understand the story-telling nuances that this project seeks to capture. Although it is an important step from simply considering melodies for emotional classification, it ultimately stops at distinct sentimental categories.

Furuya, M., et al highlighted and addressed the lack of classification methods based on lyrics in the field of music therapy [3]. While existing studies primarily classify music based on melody and tempo, this study proposes a novel approach to music classification using emotional words found in lyrics. The aim is to develop a method that can support music therapy by categorizing music based on its emotional content. The proposed method creates vector features based on the frequency of words matched with each particular emotion. Then, clustering technique is used to classify songs with similar vectors. The paper introduces a new music classification method that emphasizes the emotional aspects of lyrics, aiming to enhance its application in music therapy. However, it fails to consider the narrative elements of lyrics and solely concentrates on emotional aspects. Our project opted for NLP instead of clustering methodology to classify emotions and extends its analysis to comprehend the narratives conveyed in each song.

6. CONCLUSIONS

One limitation is the effectiveness of the machine learning model for song selection based on lyrics, which is heavily dependent on the quality and availability of the data. Limited or biased datasets may lead to inaccurate recommendations and classifications. To enhance the application's practicality, the model underwent training with a finite dataset. However, this approach is limited since additional data would undoubtedly improve its effectiveness.

Further, Analyzing lyrics for emotional content and narrative elements can be challenging due to the complexity and nuances of language. The use of keyword matching, although effective for surface-level analysis, presents shortcomings when trying to capture subtle emotional cues and interpret the deeper meanings conveyed in lyrics. Also, the model is trained entirely in English without consideration of other languages. Emotions and storytelling elements expressed in lyrics can vary significantly across different languages and cultural contexts, impacting the model's effectiveness and generalizability.

An additional constraint lies within Firebase storage [11]. It is unfeasible to pre-analyze and store all songs that exist in the database. It will also take some time for newly released songs to be included. Developer bias may occur, where the songs added to the database are predominantly mainstream or within the developer's personal knowledge base, resulting in a limited scope of information. As a result, less popular and more obscure genres may be disregarded. This constraint impacts the recommendations since they solely rely on the database content. There's a possibility that songs matching the user's description might be available but haven't been included in the database.

In summary, this study represents a step towards harnessing technological advancements to augment music therapy based on lyrical analysis. Through the integration of Firebase for backend infrastructure, Flutter for frontend interface development, natural language processing and machine learning algorithms for lyric-based song selection, our endeavor aspires to make tailored music recommendations more accessible.

Nevertheless, continual endeavors to enrich the database, enhance model precision, and integrate user input are imperative for the sustained success of the initiative.

Looking ahead, we hope to embrace innovation and be responsive to user requirements. We aspire to establish a platform that not only enriches lives through music but also serves as a beacon of technological advancement in the domain of therapy and wellness.

REFERENCES

[1] Xu, Liang, et al. "Using machine learning analysis to interpret the relationship between music emotion and lyric features." PeerJ Computer Science 7 (2021): e785.

[2] Kim, Minho, and Hyuk-Chul Kwon. "Lyrics-based emotion classification using feature selection by partial syntactic analysis." 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence. IEEE, 2011.

[3] Furuya, Mizuki, Hung-Hsuan Huang, and Kyoji Kawagoe. "Music classification method based on lyrics for music therapy." Proceedings of the 18th International Database Engineering & Applications Symposium. 2014.

[4] Chowdhary, KR1442, and K. R. Chowdhary. "Natural language processing." Fundamentals of artificial intelligence (2020): 603-649.

[5] Watson, Bruce William. "The performance of single-keyword and multiple-keyword pattern matching algorithms." (1994).

[6] Tashildar, Aakanksha, et al. "Application development using flutter." International Research Journal of Modernization in Engineering Technology and Science 2.8 (2020): 1262-1266.

[7] Fell, Michael, and Caroline Sporleder. "Lyrics-based analysis and classification of music." Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers. 2014.

[8] Vall, Andreu, et al. "Order, context and popularity bias in next-song recommendations." International Journal of Multimedia Information Retrieval 8 (2019): 101-113.

[9] Pennebaker, James W., Martha E. Francis, and Roger J. Booth. "Linguistic inquiry and word count: LIWC 2001." Mahway: Lawrence Erlbaum Associates 71.2001 (2001): 2001.

[10] Fang, Jiakun, et al. "Discourse Analysis of Lyric and Lyric-Based Classification of Music." ISMIR. 2017.

[11] Khawas, Chunnu, and Pritam Shah. "Application of firebase in android app development-a study." International Journal of Computer Applications 179.46 (2018): 49-53.

[12] Panda, Renato, et al. "How does the spotify api compare to the music emotion recognition state-of-the-art?." 18th Sound and Music Computing Conference (SMC 2021). Axea sas/SMC Network, 2021.

[13] Ofoeda, Joshua, Richard Boateng, and John Effah. "Application programming interface (API) research: A review of the past to inform the future." International Journal of Enterprise Information Systems (IJEIS) 15.3 (2019): 76-95.

[14] Bassett, Lindsay. Introduction to JavaScript object notation: a to-the-point guide to JSON. " O'Reilly Media, Inc.", 2015.

[15] Biau, Gérard. "Analysis of a random forests model." The Journal of Machine Learning Research 13 (2012): 1063-1095.