Advancing Music Therapy with Machine Learning: A Comprehensive Analysis of the MxMH Data

Applied Machine Learning

2024/06/16

陳姿妤 會計四 109303073

Table of content

I. Introduction

A. Dataset

The MxMH dataset, titled "mxmh_survey_results.csv" and available on Kaggle.com, was provided by Catherine Rasgaitis. It focuses on the study of music therapy (MT), a recognized evidence-based practice that utilizes music to enhance individual well-being. Specifically, MT leverages the power of music to positively affect stress levels, mood, and overall mental health, often by stimulating the release of "happy" hormones such as oxytocin.

Unlike traditional therapies, music therapy employs a diverse range of music genres, which can vary significantly from one organization to another. The MxMH dataset seeks to explore potential correlations between an individual's music preferences and their self-reported mental health outcomes. The insights gleaned from this analysis could potentially refine the application of music therapy, offering a tailored approach that considers individual musical tastes, or simply provide intriguing perspectives on the interplay between music and mental health.

B. Objectives and scope

The primary objective of this project is to develop machine learning models, including both classification and regression, to analyze the MxMH dataset titled "mxmh_survey_results.csv" provided by Catherine Rasgaitis on Kaggle.com. By leveraging various machine learning techniques, we aim to uncover patterns and insights that could refine the application of music therapy. Specifically, the analysis could offer a tailored approach to music therapy, considering individual musical tastes, or provide new perspectives on the relationship between music and mental health. This project aspires to contribute to the broader understanding of how personalized music interventions might optimize mental health benefits.

C. Data Explorations

TABLE 1 shows the complete description of the dataset fields:

Table 1

Complete description of the dataset fields

Variable Name	**Description**
Timestamp	Date and time when form was submitted
Age	Respondent's age
Primary streaming service	Respondent's primary streaming service

Hours per day	Number of hours the respondent listens to music per day
While working	Does the respondent listen to music while studying/working?
Instrumentalist	Does the respondent play an instrument regularly?
Composer	Does the respondent compose music?
Fav genre	Respondent's favorite or top genre
Exploratory	Does the respondent actively explore new artists/genres?
Foreign languages	Does the respondent regularly listen to music with lyrics in a language they are not fluent in?
BPM	Beats per minute of favorite genre
Frequency [Classical]	How frequently the respondent listens to classical music
Frequency [Country]	How frequently the respondent listens to country music
Frequency [EDM]	How frequently the respondent listens to EDM music
Frequency [Folk]	How frequently the respondent listens to folk music
Frequency [Gospel]	How frequently the respondent listens to Gospel music
Frequency [Hip hop]	How frequently the respondent listens to hip hop music

Frequency [Jazz]	How frequently the respondent listens to jazz music
Frequency [K pop]	How frequently the respondent listens to K pop music
Frequency [Latin]	How frequently the respondent listens to Latin music
Frequency [Lofi]	How frequently the respondent listens to lofi music
Frequency [Metal]	How frequently the respondent listens to metal music
Frequency [Pop]	How frequently the respondent listens to pop music
Frequency [R&B]	How frequently the respondent listens to R&B music
Frequency [Rap]	How frequently the respondent listens to rap music
Frequency [Rock]	How frequently the respondent listens to rock music
Frequency [Video game music]	How frequently the respondent listens to video game music
Anxiety	Self-reported anxiety, on a scale of 0-10
Depression	Self-reported depression, on a scale of 0-10
Insomnia	Self-reported insomnia, on a scale of 0-10
OCD	Self-reported OCD, on a scale of 0-10
Music effects	Does music improve/worsen respondent's mental health conditions?
Permissions	Permissions to publicize data

II. Methodology

A. Data Pre-processing

1. Data cleaning

The first stage of data preprocessing is data cleaning. After downloading the dataset, we begin by addressing columns with excessive missing values. For instance, the 'BPM' (Beats Per Minute) column, which contains 107 missing entries, is dropped due to the high number of NaNs. Subsequently, we remove any rows that still contain missing values to ensure the dataset's completeness. Following these data cleaning steps, we are left with 718 data points ready for further analysis.

2. Introducing new columns

(1) Adding the 'fav_frequently' Column We have introduced a new binary column, 'fav_frequently', to indicate the frequency at which respondents listen to their favorite genre of music. This column assigns a value of 1 to individuals who report listening to their favorite genre "very frequently," and a value of 0 to those who do not. This feature, derived from their self-reported listening habits, is intended to capture the intensity of their engagement with preferred music styles.

(2) Adding the 'Is_Loud_Music' Column We also added a new column, 'Is_Loud_Music', to classify the respondents' favorite music genre as either loud or not loud. Following the definition of loud genres in our dataset, which includes Rock, Metal, Hip hop, Rap, EDM, K pop, Pop, and Video game music, this column is set to 1 for genres considered loud, and 0 for genres not categorized as loud (e.g., Classical, Jazz, Folk). This distinction helps us understand the types of auditory environments preferred by the respondents and may provide valuable insights for studies investigating the relationship between music volume, type, and psychological effects and behaviors.

3. Simplifying the Dataset by Removing Irrelevant Features

We have strategically removed certain columns that are less relevant to our analysis objectives. Specifically, the Timestamp, Primary streaming service, Permissions, and Fav genre columns have been dropped from the dataset.

The rationale behind this decision is as follows:

- Timestamp: This column, which records the date and time when each form was submitted, is not necessary for our analysis as it does not contribute to understanding music listening behaviors or mental health outcomes.
- Primary streaming service: While it might provide context about where respondents listen to their music, it does not directly impact their psychological responses or preferences, which are our primary focus.

- Permissions: This column, likely related to consent for using the data, is irrelevant for the statistical analysis and modeling.
- Fav genre: Although it offers insight into music preferences, our model will concentrate on more quantifiable metrics such as listening frequency and effects on mental health, which can be analyzed without this specific genre information.

4. Standardization

For the column 'Age', we transform values to a given range from 0 to 1.

For categorical fields, we converting them into numerical values:

- Conversion of Frequency Responses: The dataset contains several columns representing the frequency at which respondents listen to various music genres, ranging from 'Never' to 'Very frequently'. We mapped these categorical responses to numerical values: "Never" to 0, "Rarely" to 1, "Sometimes" to 2, and "Very frequently" to 3. This conversion was applied across all relevant columns such as 'Frequency [Classical]', 'Frequency [Country]', 'Frequency [EDM]', etc.
- Binary Conversion of Behavioral Attributes: Additionally, we standardized responses in columns related to behavioral attributes, such as 'While working', 'Instrumentalist', 'Composer', 'Foreign languages', and 'Exploratory', converting "No" to 0 and "Yes" to 1.

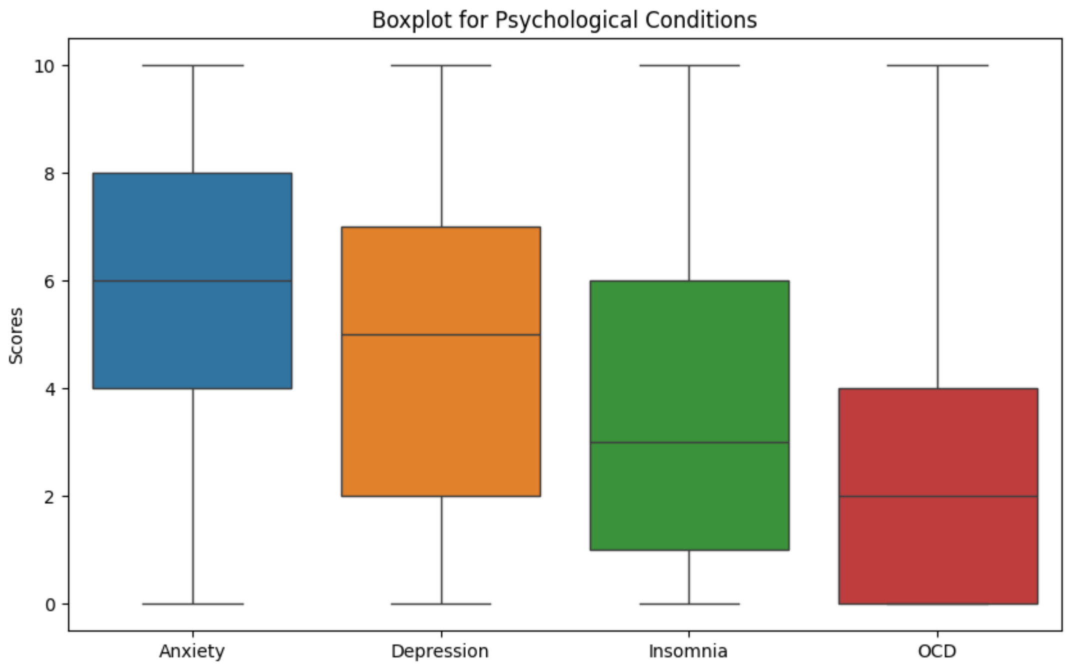
- Mapping Music Effects on Mental Health: Finally, responses in the 'Music effects' column, which describe the impact of music on respondents' mental health conditions, were converted from textual to numerical values—0 for "Worsen", 1 for "No effect", and 2 for "Improve".

B. Exploratory visualization

The data shows that among the respondents, anxiety is the most severe mental condition on average, followed by depression. Insomnia and OCD come next in terms of severity.

Figure 1

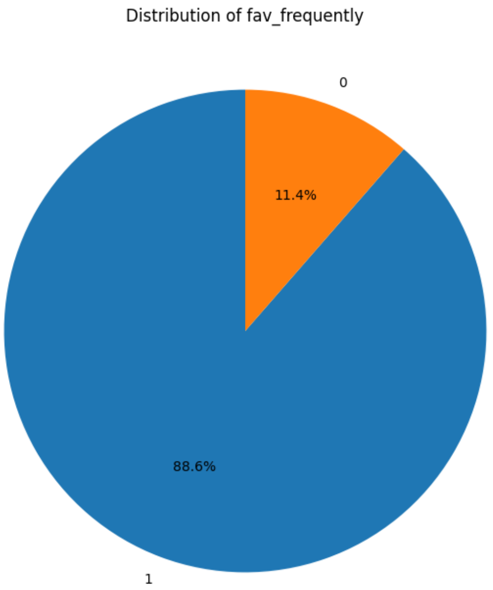
Box-plot for Psychological Conditions



The majority of participants frequently listen to their preferred music genre, with the proportion reaching as high as 88.6%.

Figure 2

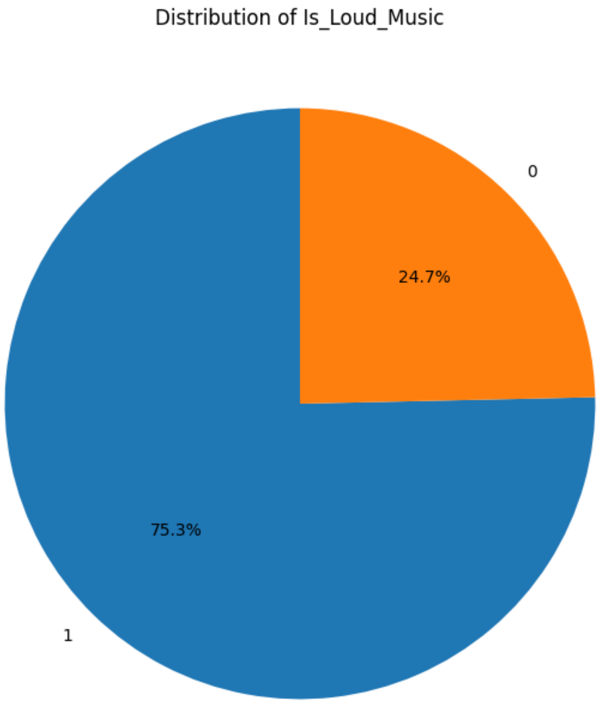
Distribution of fav_frequently



Approximately 75% of the respondents prefer listening to relatively calm music genres.

Figure 3

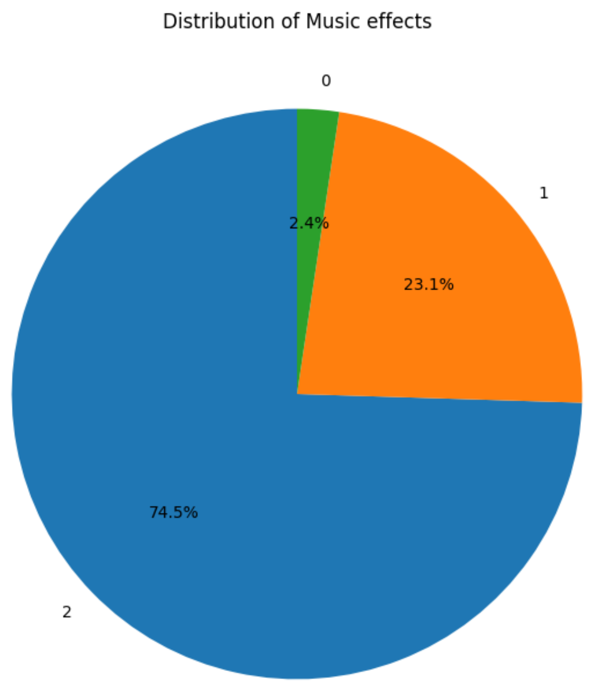
Distribution of Is_Loud_Music



As many as 74.5% of people believe that music improves their mental condition, 23.1% reported no effect, and 2.4% stated it worsened their condition.

Figure 4

Distribution of music effects



C. Modeling (ML)/ Model Comparison

For all models, we set the test size to 0.3, meaning 30% of the data is used for testing and 70% for training.

1. Classification

We train the data using three classification models: KNN, logistic regression, and SVM. For the KNN model, we consider two weight options, uniform function and distance function, and seek the best parameters by testing k neighbors in the range from 1 to 30. The TABLE 2 displays the best parameters based on four evaluation metrics. We selected 11 neighbors used and uniform function for the KNN model due to its superior accuracy and recall rate.

Table 2

Comparison of best parameters testing by different evaluation metrics

Metrics	Best Score	Best Parameters
Accuracy	0.749335	{'n_neighbors': 11, 'weights': 'uniform'}
Recall	0.749335	{'n_neighbors': 11, 'weights': 'uniform'}
Precision	0.691274	{'n_neighbors': 12, 'weights': 'distance'}
F1	0.677388	{'n_neighbors': 10, 'weights': 'distance'}

For logistic regression and SVM, we directly apply the models and assess their accuracy, precision, and F1 score. The following table display the comparison of performance of the 3 models(KNN with best parameters, Logistic regression, SVM.

Table 3

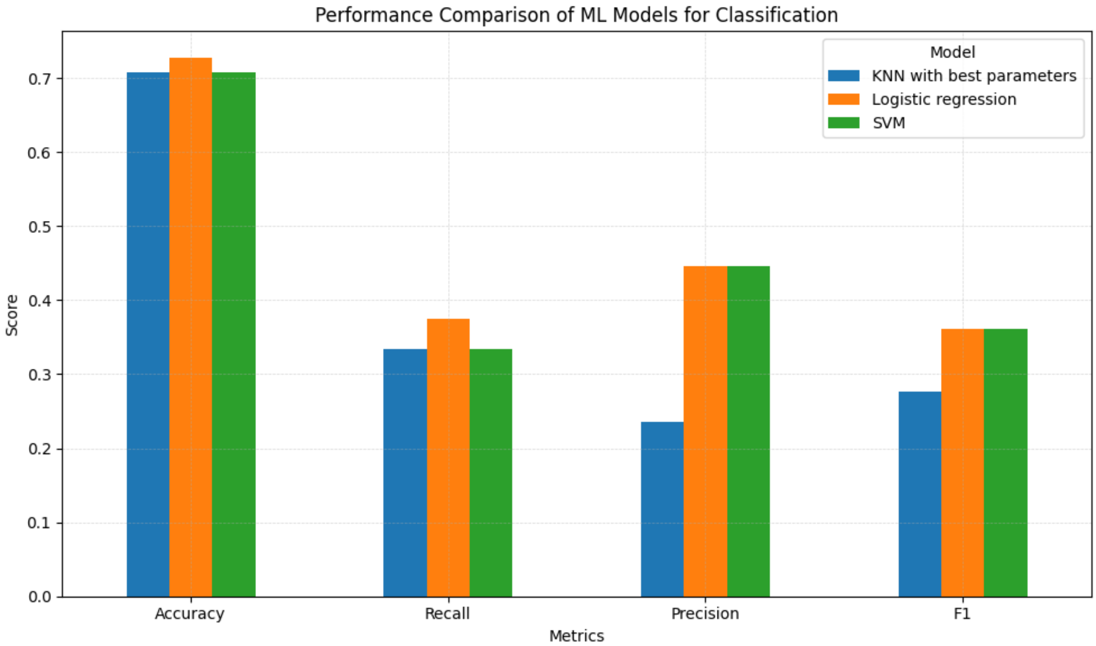
Comparison of classification models performance

Model	Accuracy	Recall	Precision	F1
KNN with best parameters	0.708333	0.333333	0.236111	0.276423
Logistic regression	0.726852	0.374164	0.445439	0.360911
SVM	0.708333	0.333333	0.445439	0.360911

From the charts provided, it is evident that the performance of the three models—KNN, SVM, and Logistic Regression—is relatively close, with Logistic Regression slightly outperforming the others. Among these, KNN is the least effective, exhibiting the lowest accuracy, recall, precision, and F1 scores. Notably, the recall, precision, and F1 scores for all models are below the satisfactory level, indicating a significant need for model improvement and refinement.

Figure 5

Performance Comparison of classification models



The confusion matrix charts further reveal that both KNN and SVM models almost completely fail to predict the "no effect"(1) and "worsen"(0) categories. This indicates significant issues with these models in distinguishing between the three categories. Logistic Regression performs somewhat better than KNN and SVM, but there is still considerable room for improvement. This analysis highlights the need for further optimization and possibly more sophisticated modeling techniques to enhance the predictive accuracy of these models across all categories.

Figure 6

Confusion Matrix for K-Nearest Neighbors

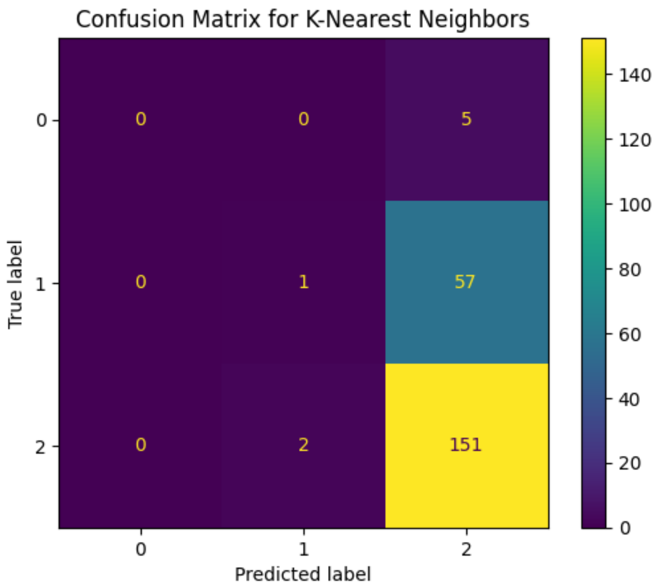


Figure 7

Confusion Matrix for logistic regression

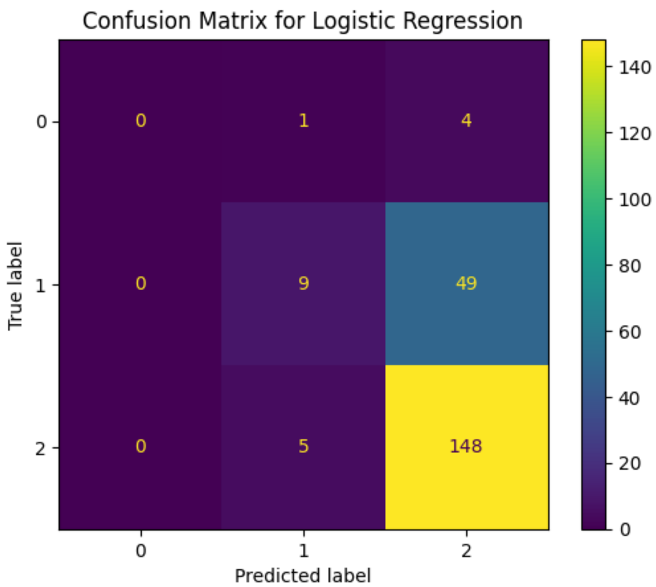
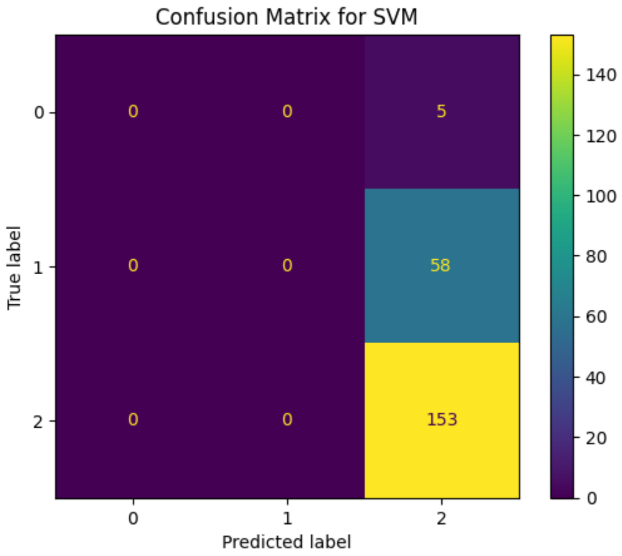


Figure 8

Confusion Matrix for SVM



2. Regularization with Linear Regression

After directly applying a regression model and finding it unsuccessful, with all feature coefficients equal to zero, I implemented various regularization techniques with linear regression. These included L1 (Lasso), L2 (Ridge), and Elastic Net. However, the performance on the test dataset has been underwhelming. Table shows a summary of the performance of these models:

Table 4

Comparison of regression models performance

Model	RMSE	MSE	R^2	MAE
Lasso	0.5037	0.2538	0.0315	0.4065
Ridge	0.5067	0.2568	0.02	0.4084

Elastic Net	0.5033	0.2533	0.0332	0.4102

These results indicate that the models are not capturing the underlying patterns effectively, suggesting potential issues such as insufficient feature relevance or data quality. The R^2 values close to zero indicate that the models explain very little of the variability in the response data, pointing to the possibility that other non-linear models or methods might be required to handle this dataset more effectively.

D. ML Model Adjustments

1. Find important variables

To identify key variables in our dataset, we employed a multi-faceted feature selection approach using three distinct methods: Univariate Selection with Select K Best, Feature Importance using the Extra Trees classifier, and Information Gain analysis. The results from the Univariate Selection using Select K Best indicated that only a few variables, such as "Depression" and "Anxiety," are significant. However, the Feature Importance analysis with the Extra Trees classifier did not reveal any additional important variables.

(1) Univariate Selection - Select K Best The following table shows the result of the Univariate Selection model.

Table 5

Result of Univariate Selection - Select K Best

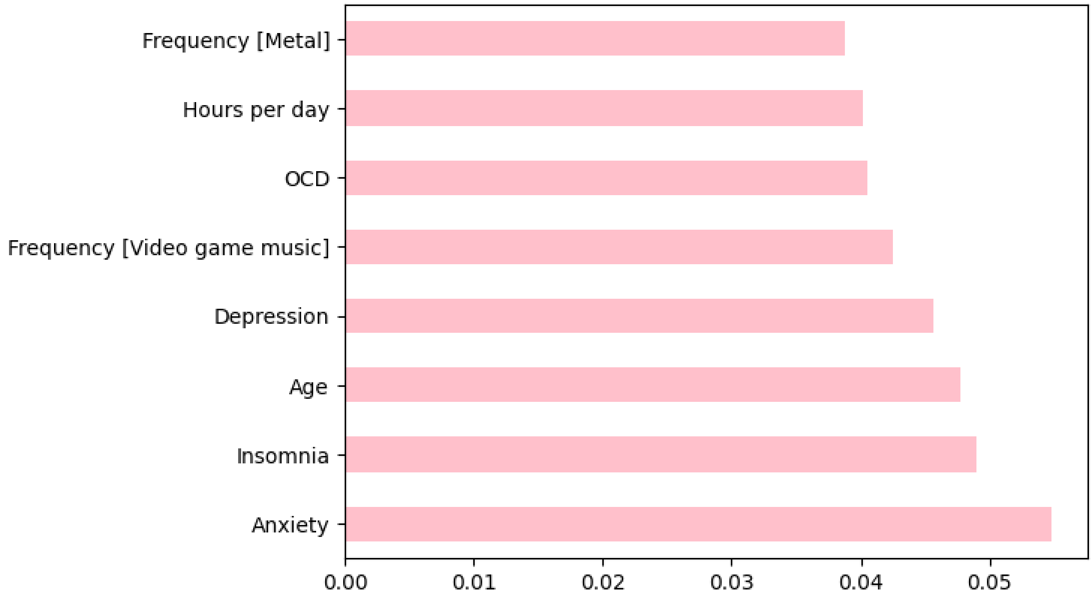
Features	**Score**
Depression	24.67519
Anxiety	21.81768
Frequency [R&B]	9.541262
Frequency [Gospel]	7.63178
OCD	5.747301
Instrumentalist	5.429344
Composer	4.776192
While working	4.716333
Exploratory	4.524447
Frequency [Hip hop]	4.40008
Frequency [Country]	4.388507
Frequency [Latin]	4.379306
Hours per day	4.181407
Frequency [Lofi]	3.943601
Frequency [K pop]	3.793949
Insomnia	3.385112
Frequency [Jazz]	2.795044

Frequency [EDM]	2.551315
Frequency [Rap]	1.927171
Frequency [Video game music]	1.92536
Frequency [Folk]	1.853213
Frequency [Rock]	1.427546
Frequency [Pop]	1.184067
Is_Loud_Music	0.955797
Frequency [Metal]	0.618566
Age	0.515851
Frequency [Classical]	0.45636
Foreign languages	0.113482
fav_frequently	0.083433

(2) Feature Importance - Extra Tree The following table shows the result of the Extra Tree model.

Figure 9

Result of Univariate Selection - Select K Best



(3) Information Gain The following table shows the result of the Information Gain model.

Table 6

Result of Univariate Selection - Select K Best

Feature	**Information Gain**
While working	0.030241
Age	0.023468
Instrumentalist	0.018644
Frequency [Country]	0.016608
Frequency [Pop]	0.014814
Frequency [EDM]	0.013447
Frequency [Hip hop]	0.013183

Exploratory	0.010803
Frequency [Classical]	0.010323
Frequency [Gospel]	0.009167
Frequency [R&B]	0.00815
Frequency [Rock]	0.003715
Frequency [Metal]	0.001572
Hours per day	0.001101
Frequency [Folk]	0.000667
Depression	0
Anxiety	0
Insomnia	0
OCD	0
fav_frequently	0
Frequency [Video game music]	0
Frequency [K pop]	0
Frequency [Rap]	0
Frequency [Lofi]	0
Frequency [Latin]	0
Frequency [Jazz]	0

Foreign languages	0
Composer	0
Is_Loud_Music	0

2. PCA & Classification

From the analysis presented in the table, it is evident that applying Principal Component Analysis (PCA) did not significantly enhance the performance of the classification models tested, which include K-Nearest Neighbors (KNN), Logistic Regression, and Support Vector Machines (SVM).

Table 7

Comparison of performance of classification model with and without PCA

Model	Accuracy	Recall	Precision	F1
KNN	0.703704	0.334723	0.347418	0.285974
KNN with PCA	0.708333	0.333333	0.236111	0.276423
Logistic regression	0.726852	0.374164	0.445439	0.360911
Logistic regression with PCA	0.726852	0.376343	0.445439	0.360911
SVM	0.708333	0.333333	0.445439	0.360911
SVM with PCA	0.708333	0.333333	0.445439	0.360911

Also, in regression model there is no significant improve or even become poorer after applying PCA. We can observe the result from following table and the R-square bar chart.

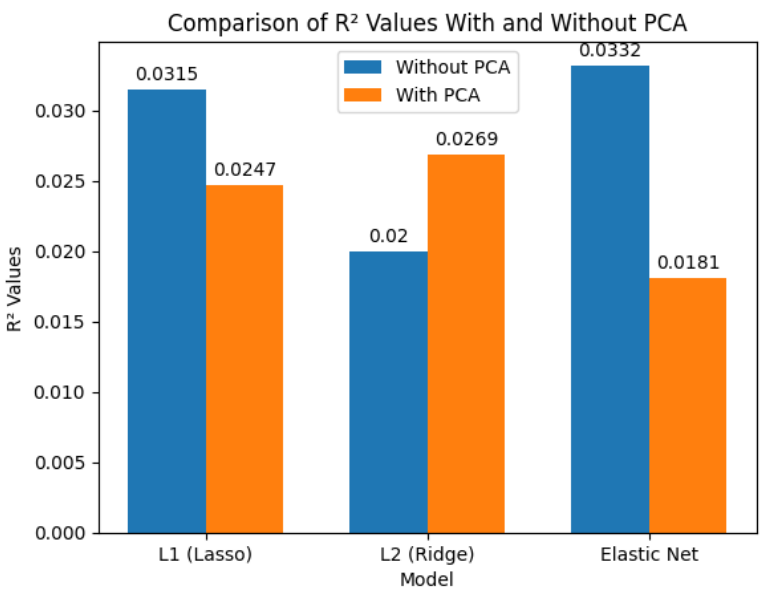
Table 8

Comparison of performance of regression model with and without PCA

Model	RMSE	MSE	R^2	MAE
Lasso	0.5055	0.2555	0.0247	0.4095
Ridge	0.5049	0.2549	0.0269	0.406
Elastic Net	0.5072	0.2573	0.0181	0.4157

Figure 10

Comparison of R-square Value of regression models with and without PCA



3. *Handling Data Imbalance with Under-Sampling, Over-Sampling, and SMOTE in Logistic Regression*

Since we were unable to identify key features and PCA did not enhance model performance, we applied under-sampling, over-sampling, and SMOTE (Synthetic

Minority Oversampling Technique) to address data imbalance. However, these methods resulted in even poorer outcomes in the Logistic Regression model with PCA.

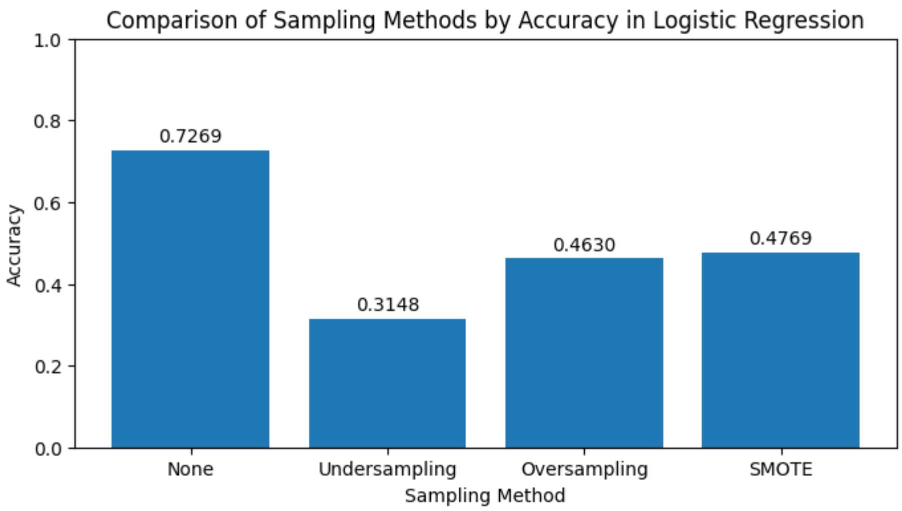
Table 9

Comparison of sampling method in Logistic Regression

Sampling Method	Accuracy	Precision	F1 Score
Under sampling	0.3148	0.34	0.2615
Oversampling	0.463	0.4219	0.3767
SMOTE	0.4769	0.3957	0.3535

Figure 11

Comparison of sampling method by accuracy in logistic regression



4. Decision tree

We applied a decision tree model because specific important features for predicting 'y' were not identifiable and PCA failed to enhance model effectiveness. According to the data in the tables, the decision tree model is unable to identify class 0, with or without PCA. Although it performs slightly better at identifying class 1 with PCA, it still falls short of being effective.

Table10

Decision tree performance without PCA

Class	Precision	Recall	F1-score	Support
0	0	0	0	5
1	0.38	0.09	0.14	58
2	0.71	0.95	0.81	153

Table 11

Decision tree performance with PCA

Class	Precision	Recall	F1-score	Support
0	0	0	0	5
1	0.4	0.14	0.21	58
2	0.72	0.93	0.81	153

III. Results

For classification, logistic regression performed relatively better among all the models tested. However, there is still significant room for improvement. The regression models, on the other hand, were a major failure. Despite efforts in feature selection, principal component analysis (PCA), and handling the imbalance of data, the performance of all models did not improve.

V. Conclusion

In the future, when considering more music features, factors such as the emotional style of the music (e.g., happy, sad, calm), the rhythm, and the volume can be included. Currently, the data is based solely on the subjective judgment of the participants to determine if there is any improvement, but it does not specify what exactly has improved (e.g., depression, anxiety, insomnia). It is essential to clearly record the specific aspects of improvement to more accurately evaluate the effectiveness of music therapy. Additionally, relying only on the participants' subjective judgment may not be precise enough; incorporating scientific measurements could yield better results.

Emotions are highly subjective, and the reasons for psychological distress vary from person to person. Different reasons may affect the effectiveness of music therapy on improving mental health. Therefore, in future data collection, it is recommended to include the reasons for psychological distress and to group the data based on these reasons, as well as the type and severity of the mental disorder. This approach may lead to more accurate predictions.

References

Music & Mental Health Survey Results

https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results/data