Analyzing the Impact of Musical Habits on Mental Health

Peter Loiselle
Department of Computer Science
CS 470
Emory University
Atlanta, GA
ploisel@emory.edu

Jason Lebov
Department of Computer Science
CS 470
Emory University
Atlanta, GA
jason.lebov@emory.edu

***Abstract*— This study investigates the impact of a person's musical listening habits on mental health, specifically looking at obsessive-compulsive disorder *(OCD)*, insomnia, anxiety, and depression. This analysis was performed using a Kaggle dataset called "Music and Mental Health Survey Results", which surveyed people on the internet regarding their musical interests, habits, and overall mental states. The data was cleaned and preprocessed accordingly, and then was fed into several machine learning models with ranging complexities, including KNN, Decision Tree, Random Forest, and a Neural Network, and evaluated using a wide range of metrics such as accuracy and ROC. The goal of this study was to uncover correlations between musical characteristics and mental health to predict both the presence of mental illness and its severity. Despite extensive analysis, this study found little evidence to support a significant relationship between an individual's musical preferences and their overall mental well-being. While this study did not yield conclusive results, further research could be performed using alternative methods along with larger, more developed, datasets to better uncover relationships. This research is important due to its potential impact on mental health services, including diagnostic capabilities and therapies. Furthermore, identifying potential correlations between music and mental disorders, such as which genres correlate to different illnesses, could revolutionize how people engage with music and other related habits.**

Keywords*— *Music, Mental Health, KNN, Decision Tree, Random Forest, Neural Network, Diagnostic, Therapy

I. INTRODUCTION

Music, throughout history, has played an integral role in society, as people listen to their favorite artists and songs daily. Music, for many, is just another part of a daily routine as people listen while they commute to work, go to the gym, make dinner, and do numerous other activities that make up their day. According to a 2014 report done by Nielsen Music titled, '2014 Nielsen Music U.S. Report', 93% of the U.S. population listen to music daily which on average occupied around 25 hours of their time each week. Additionally, this report found that 75% of people would rather spend their time listening to music over other forms of entertainment such as watching TV [1].

While music is often regarded as a positive stimulant, used to relax and relieve stress, given its extreme prevalence in people's everyday lives, it is important to understand if any negative effects exist. Specifically, if there are any potential negative mental health impacts of music listening.

Mental illness in the United States, and around the world, has become a significant issue throughout society, affecting the lives of a large portion of the population. A 2021 study conducted by the National Alliance on Mental Illness found that in the US alone, 22.8% of adults and 16.5% of youth experienced some form of mental illness [2].

Given the high rates of both music listening and people affected with mental disorders, this study aims to investigate the potential correlation between an individual's musical habits and mental health, specifically looking at obsessive-compulsive disorder (OCD), insomnia, anxiety, and depression. Understanding potential correlations could change the way people listen to music, as well as change current approaches to mental illness-related diagnostic procedures and therapies.

II. RELATED WORK

Current reliable and well-developed research investigating the relationship between music preferences and mental disorders is limited, but some studies touch on the interactions between music and mental health in other, more general ways. Although this prior research and work exists, there is still room for further analysis into the specifics of their relationship. Specifically, no studies have been performed to identify musical habits as warning flags for mental disorders. That said, there have been studies about music's relationship with people who are already known to have a given disorder. For example, A 2021 study from Frontiers in Psychology determined that music does indeed have a positive effect on the mental health of patients suffering from mental disorders [3]. However, the study did not account for different genres of music, and it incorporated other forms of treatment such as dance therapy, which made the results less controlled. Another study from 2015 determined that a large portion of patients decrease the frequency with which they listen to music when afflicted with a mental disorder [4]. This study also found that patients who listen to rock, metal, pop, or folk music tend to keep these as their favorite genres when afflicted with a disorder, while patients who listen to other genres often switch to a new genre. However, the reasoning behind this change in type and frequency after onset remains unknown, and there was no significant insight into genre preference and whether it helped or harmed a patient's mental state. By contrast, our study will focus directly on using musical habits to predict exactly what a person may be suffering from. The effects of this study will hopefully shed further light on the impacts music may have, adding more detail to this field of study and even potentially adding new insights such as genre-specific effects on mental health.

III. DESCRIPTION OF DATA

The dataset used in this study was sourced from Kaggle.com and was collected and updated in 2024 as part of a different study analyzing the impacts of music therapy on an individual's mental well-being. Named ‘Music & Mental Health Survey Results’, this dataset consists of 736 respondents and 33 attributes. These attributes collected information regarding musical characteristics/habits and mental state of each respondent. The mental illnesses it collected data on are obsessive-compulsive disorder (OCD), insomnia, anxiety, and depression.

Specifically, attributes relating to the respondents' listening habits, collected information regarding the respondent’s favorite genre, frequency, and circumstance of music listening, as well as other relevant information such as total time listening per day and circumstances of listening. Additional attributes offer insight into each respondent’s mental state, highlighting any type of mental disorder they are experiencing and its severity (scale 0-10). Basic respondent information is also present in the dataset to contextualize each person’s identity.

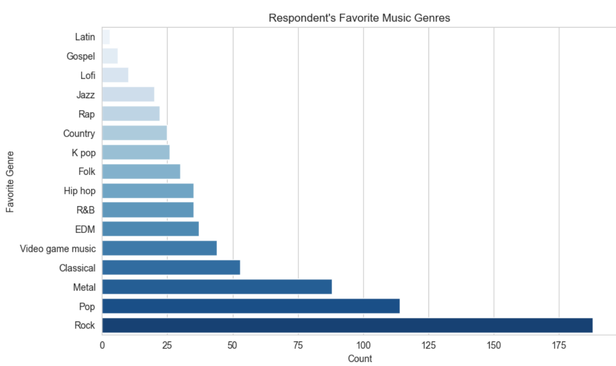

Fig 1. Distribution of Respondent’s Favorite Genres

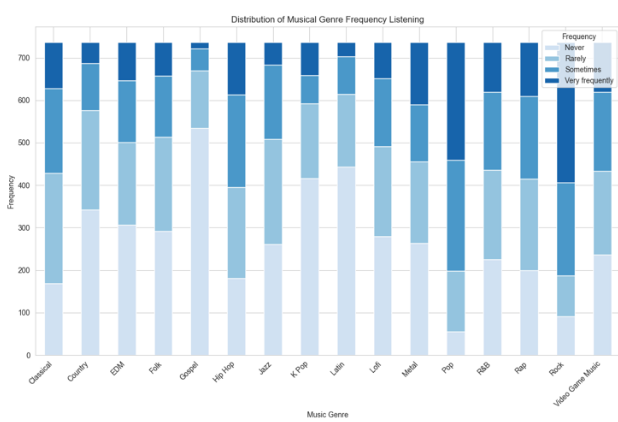

Fig 2. Distribution of Respondent’s Music Listening Frequency Across Genres

IV. METHODOLOGY & SYSTEM DESIGN

For this study, we used the Python programming language to perform our analysis and specifically utilized Pandas and scikit-learn packages.

Our overall approach to perform this analysis was to break up the overall base dataset, from Kaggle, based on the four mental illnesses it had information on: Anxiety, Depression, OCD, and Insomnia. This resulted in four sub-datasets, each having a single target column for the specific mental illness. Each of these datasets were then used for training and testing with four different machine learning models.

Additionally, during preprocessing, we transformed the label feature from its original severity scale of 0-10, into two different types of classification tasks: Multi-Class Classification and Binary Class Classification. For the Multi-Class datasets, we transformed the feature into four categories: Asymptomatic (0.0-1.0), Mild (1.5-4.0), Moderate (4.5-7.0), or Severe (7.5-10.0) to understand the severity with which the respondent suffered. For the Binary datasets, we converted the values to No (0-4.9) or Yes (5-10), for whether or not the respondent suffers at all from the illness.

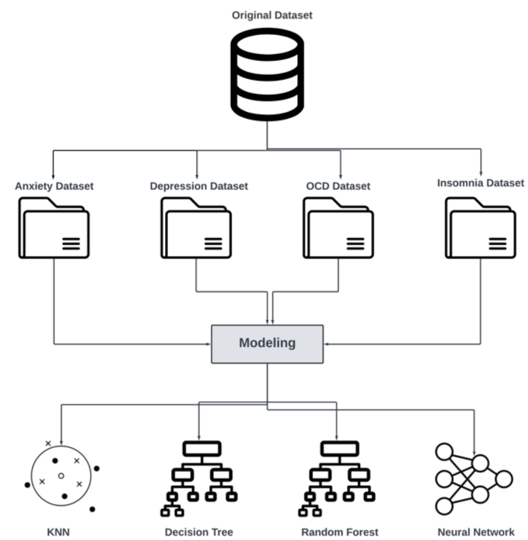

Fig 3. Overall System Design for Both Multi-Class & Binary Classification Tasks

V. DATA PREPROCESSING

The first step in designing our system involved preprocessing our data. This process included multiple steps such as feature cleaning & extrapolation, feature selection, dataset creation, and feature scaling.

A. Feature Cleaning & Extrapolation

In terms of feature cleaning the original dataset, steps were taken to remove any unwanted columns or samples that we felt would impact our overall results. Preliminary analysis was conducted on the dataset to identify features that we deemed as irrelevant, and that we did not want to keep in our final datasets. These features included the timestamp of when the survey was taken (‘Timestamp’), the respondent consenting to their survey results being used in a study (‘Permissions’), the BPM of the respondent’s favorite genre (‘BPM’), and whether listening to music improved the respondent’s mental health (‘Music Effects’). ‘Timestamp’ and ‘Permissions’ were immediately dropped because they provided no context to the user’s musical listening habits, and therefore would not be correlated to our target feature. ‘BPM’ was dropped because it had many missing values, so we felt it was best to drop it to avoid adding random or skewed values to our data by filling in missing values with other, calculated values. Lastly, ‘Music Effects’ was dropped because it did not relate to our research question of predicting mental illness along with the fact that it had a very unbalanced response rate with over 85% of respondents saying that listening to music improved their mental health or had no effect. However, this feature could be used in further research to investigate other impacts of musical habits. Additionally, samples with NaN values were also dropped from the dataset, which accounted for only eight samples after dropping the 'BPM' feature which had over 100 missing values.

Once basic cleaning was performed on the data, extrapolation was done to convert all the categorical values to numeric. Converting all categorical variables to numeric is valuable as it allows for increased model performance as well as allows for more efficient, and simplistic data preprocessing.

Specifically, One-Hot Encoding and Label Encoder from scikit-learn were used to achieve this. The features that were One-Hot Encoded include 'Primary Streaming Service' and 'Fav Genre'. These features were One-Hot Encoded rather than Label Encoded because they had low cardinality and therefore, were able to be converted to numeric without greatly increasing our dimensionality. For features with higher cardinality, Label Encoder was used, and these columns included, 'While working', 'Instrumentalist', 'Composer', 'Exploratory', and 'Foreign Languages'. Also, the 'Frequency [__]' features, that accounted for the frequency of listening for each genre, were also Label Encoded into 0 (*Never*), 1 (*Rarely*), 2 (*Sometimes*), and 3 (*Very Frequently*)

Lastly, the target features (Anxiety, Depression, OCD, & Insomnia) were also engineered to convert the values for the original 0-10 scale to categorical variables, and then back to more defined numeric values using Label Encoder. This was done both for the Multi-Class and Binary Class part of our study. For the Multi-Class engineering, the data was converted into four categories: Asymptomatic (*0.0-1.0*), Mild (*1.5-4.0*), Moderate (*4.5-7.0*), and Severe (*7.5-10.0*). These values were then Label Encoded back to numeric values where 0 = Asymptomatic, 1 = Mild, 2 = Moderate, and 3 = Severe. For the Binary Class engineering, the values were converted to two categories: No (*0-4.9*) & Yes (*5-10*). These were then also converted back to numeric values where 0 = No and 1 = Yes.

B. Feature Selection

For the feature selection aspect of this study, Pearson Correlation Matrices were used to determine which features were highly correlated to each other as well as which features were highly uncorrelated to the target feature. Several correlation matrices were created for each of our datasets, across each mental illness, and for each classification type (Multi & Binary). Features that were highly correlated to each other (*above determined Delta threshold*) as well as features that were not correlated to the target feature (*below determined Gamma threshold*) were dropped accordingly. After performing this analysis, the feature that was dropped based on the Delta threshold was 'Frequency of Hip-Hop' which was highly correlated to the 'Frequency of Rap' feature. No features were dropped based on the Gamma threshold.

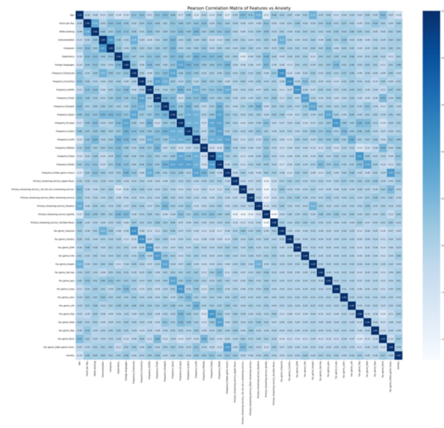

Fig 4. Pearson Correlation Matrix for the Anxiety Dataset (Results Were the Same Across All Datasets)

C. Dataset Creation

Next, using the fully engineered base dataset, the sub-datasets were created, each split up based on mental illness: Anxiety, Depression, OCD, & Insomnia. All datasets (both for Multi-Class and Binary) had the exact shape, with the same number of samples and features.

An oversampling technique called SMOTE was applied to the Binary Class training datasets because these suffered from a severe class imbalance problem. SMOTE generates more samples of the minority class, creating a uniform distribution of the target class.

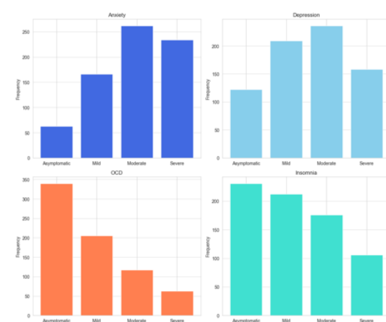

Fig 5. Target Class Distribution for Multi-Class Datasets

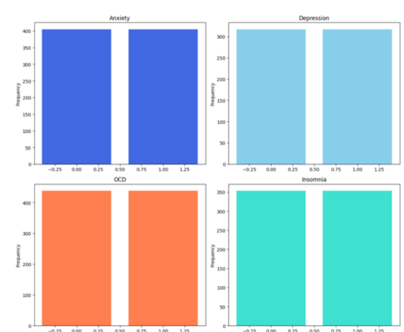

Fig 6. Target Class Distribution for Binary-Class Datasets after Applying SMOTE Oversampling Technique

D. Feature Scaling

Lastly, after wrapping up feature selection and creating the Train-Test Splits of all created sub-datasets, Min Max Scaling was then performed on all features (*given all are numeric*) on the training data. This was done to ensure that all values had equal weight when passed into our models. Also, this scaling was a necessity to normalize the data regardless of magnitude or units by scaling all values from the range of 0-1. This ensured that every feature was considered equally and would allow the implemented models to learn the data best, given all values are now uniform.

VI. MODEL SELECTION & CREATION

In terms of model selection for this study, we chose to focus on a variety of models, all ranging in complexity: KNN, Decision Tree, Random Forest, and Neural Network. We felt that by selecting these models, we would be able to get a better conclusion to our research question, as well as be able to identify which algorithm worked best for each mental illness. Furthermore, we chose these models because they are known for being able to handle smaller sized datasets, which we have, as well as be able to prevent overfitting. Specifically, KNN is good at this given its overall simplicity [5] while Random Forest and Neural Networks are also good at this because they are both capable of creating generalizations from smaller datasets [6].

A. Model Construction

Our four models were built identically across the four datasets (for both Mutli-Cass and Binary), with the only difference being the hyperparameters and the target feature. We created these models using scikit-learn's built-in KNN, Decision Tree, Random Forest, and MLPClassifier (Neural Network) models. We heavily utilized pandas for data manipulation, such as creating feature and target value splits, and to easily run our datasets through our models.

B. Hyperparameter Tuning

When creating our models, we also utilized scikit-learn's GridSearchCV function to find the optimal hyperparameters for each of our models. Given we have four different models and four different datasets for each model, we ran Grid Search 16 times to find the best parameters for each. This was done to ensure we were optimizing our model's performance and accuracy. Based on each model's optimal output, we constructed each model based on its specific parameters. To do this efficiently using GridSearchCV, we ran each model with varying parameter grids to see how our results were affected by each. For example, when creating our KNN models, we used Grid Search to find the optimal number of neighbors as well as the best weight for each model. For our Decision Trees, we found the best criterion, max depth, minimum size leaf samples, and minimum size of samples to split on. For the Random Forests, we tested for the optimal number of estimators/trees, max depth, minimum number of leaf samples, and minimum number of samples to split on. Lastly, for the Neural Networks, we tested for the best hidden layer sizes, activation function, solver, learning rate, and alpha value. This analysis resulted in each model having different optimal hyperparameters based on the dataset (*mental illness*) it was being trained and tested on.

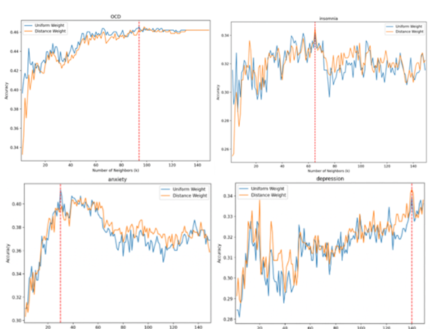

Fig 7. Best Hyperparameters for KNN Models (N-Val & Weight)

VII. RESULTS

After performing analysis using the sixteen models trained and tested on the Mutli-Class datasets with their optimal hyperparameters, sixteen unique accuracy values were found. As seen in *Table 1* and *Fig 8.* below, all sixteen accuracy values across all datasets and models, although on the lower end, were still higher than random chance (25%). Additionally, it can be seen that for each of the chosen models, there are ranging accuracy scores observed, with the bold values being the best model for each dataset/mental illness.

	Depression	Anxiety	OCD	Insomnia
KNN	0.290	**0.393**	0.500	0.310
Decision Tree	**0.343**	0.369	0.424	**0.345**
Random Forest	0.303	0.352	**0.510**	0.269
Neural Network	0.324	0.269	0.331	0.262

Table 1. Accuracy Values for all Multi-Class Models

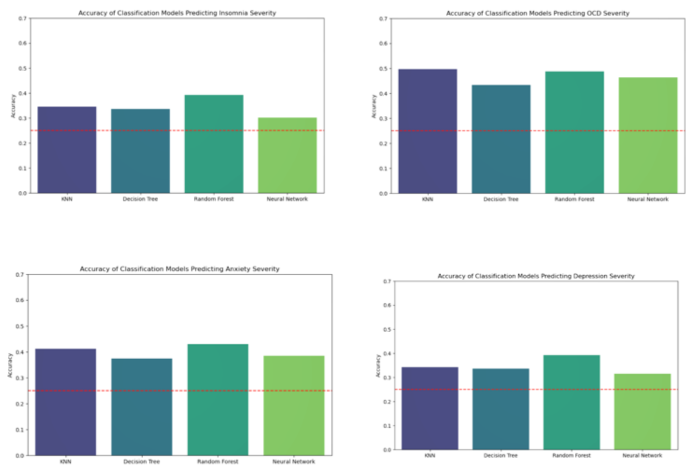

Fig 8. Bar Chart of Accuracy Values for all Multi-Class Models

Based on these above results, there are clearly models that are better fit for each dataset and illness. While our best accuracies for correlations with anxiety, depression, and insomnia were higher than random chance, they were only marginally higher. However, OCD performed better overall as it had over 50% accuracy, but upon further investigation, we determined this may have been due to a slight class imbalance in this dataset. In summary, we determined that there is no strong correlation between a person's musical habits/characteristics and the severity with which they suffer from mental disorders.

Our results for the Binary data, as seen in *Table* 2, ultimately led to similar findings. While some accuracy values were higher than random chance (50%), there was not a large enough difference to claim that you could accurately predict whether a person suffers from a mental disorder based on musical habits/characteristics.

	Depression	Anxiety	OCD	Insomnia
KNN	**0.586**	0.469	0.331	0.497
Decision Tree	0.538	**0.579**	0.641	0.517
Random Forest	0.283	0.380	**0.717**	0.280
Neural Network	0.580	0.580	0.676	**0.538**

Table 2. Accuracy Values for all Binary Models

Furthermore, a closer analysis of the binary results showed that our findings were about the same as random guessing, which we could already conclude from our accuracy scores. Our ROC cures, shown in *Fig* 9, sat right around the random chance baseline for all of our datasets and all of our models, with a few being slightly elevated: mainly for OCD and Depression.

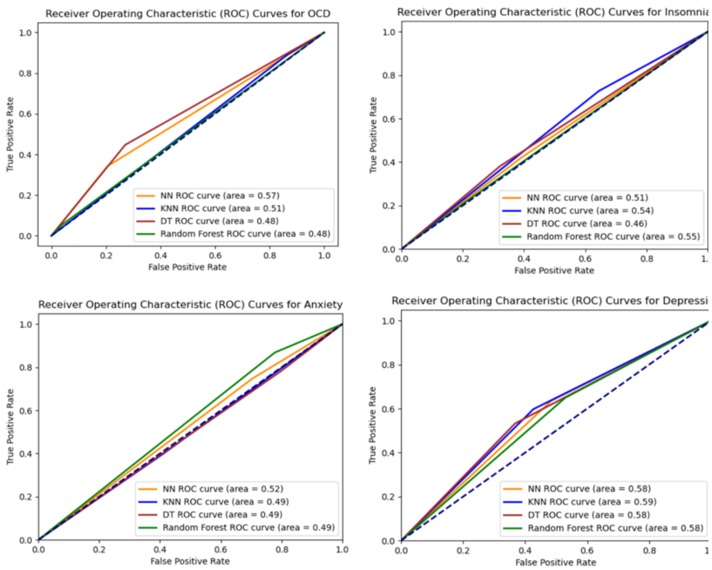

Fig 9. ROC Curves of All Binary Models for Each Mental Disorder Investigated

Our initial assumption was that our low accuracies & ROC scores were due to overfitting for the Binary data, especially given the small sample size of all of the datasets. However, after performing analysis into feature importance for each of the models, we concluded that no feature for any model or dataset had any significant impact on our target variables. As seen in *Fig 10-13,* all importance scores were lower than 0.15, indicating that there was no strong correlation between any of the features and our intended output. This could potentially also be a result of the small datasets, given our models may not have enough data to determine the importance of each feature.

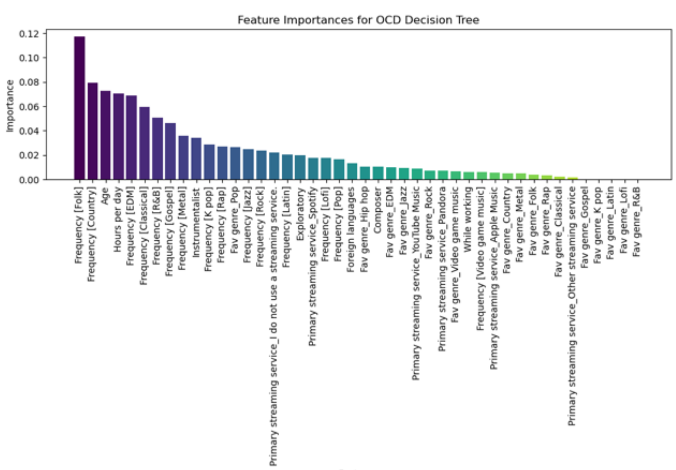

Fig 10. Feature Importance Graph for Binary OCD DT Model

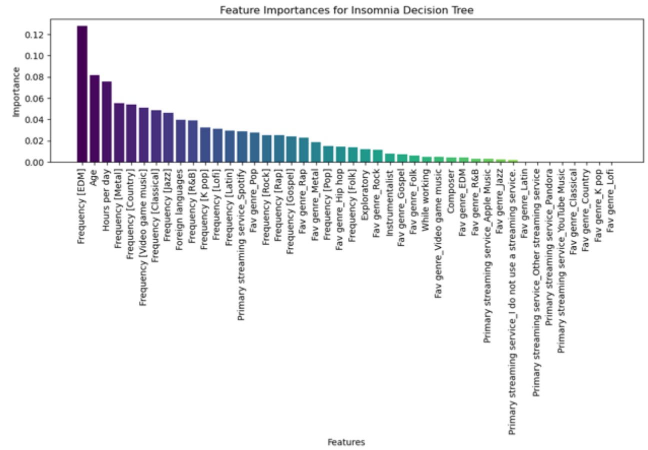

Fig 11. Feature Importance Graph for Binary Insomnia DT Model

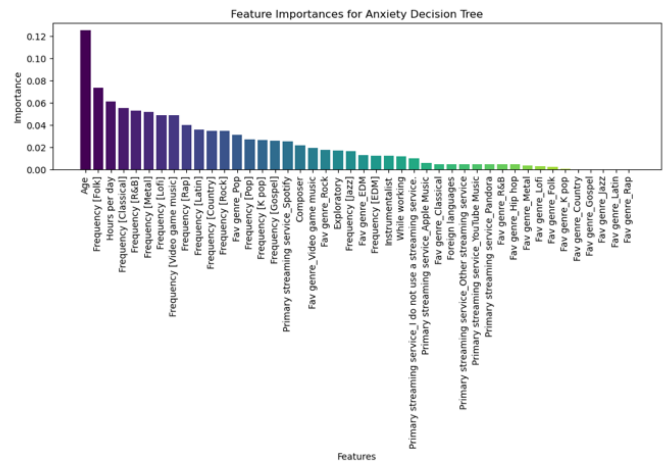

Fig 12. Feature Importance Graph for Binary Anxiety DT Model

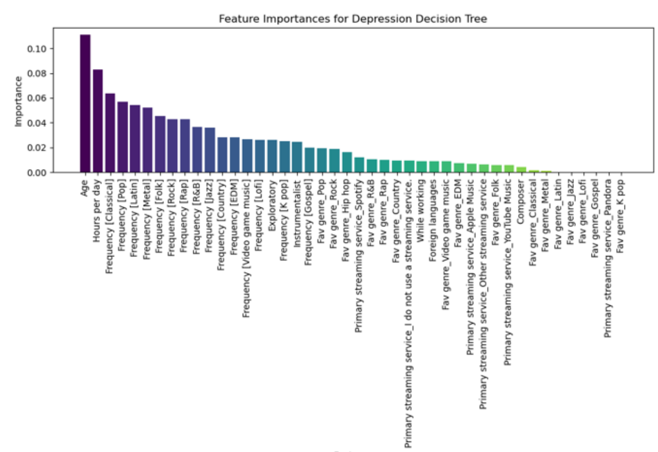

Fig 13. Feature Importance Graphs for Binary Depression DT Model

VIII. DISCUSSION

1. Can machine learning be used to predict mental illness severity based on music preferences?

No - Our results indicate that there is no notable correlation between a person's music habits and their mental state. Genre, frequency, and musical involvement all appear to play little to no role in how badly a person suffers from a given mental disorder.

This is clear from the accuracy scores produced by all four models for each examined illness.

2. *Can machine learning be used to predict the presence of mental illness based on music preferences?*

No - Based on our results for the Binary Class datasets, using music preferences as an indicator of mental health is no more effective than randomly guessing whether a person suffers from a mental disorder. This is illustrated by accuracy scores near 0.5 and ROC scores also around 0.5 across the board.

3. *Which Features Appear to be the most important when it comes to predicting Mental Illness?*

There are no features that consistently prove to be important in determining a person's mental state. In the feature importance graphs, *Fig 10-13*, no feature had an importance greater than 0.15 for its predictive power towards the target. Thus indicating that no features have a strong correlation with mental health at all. The "Age" feature was consistently among the most important features, indicating that age could have a correlation with mental health, but its score was always too low to indicate any meaningful insight.

Overall, based on similar results across four different models and eight different datasets (Multi-Class & Binary), and because of consistently low feature importance values, we ultimately determined that a person's musical habits and characteristics have little to no correlation with their mental health, either when looking at the severity of a mental disorder or simply the existence of it. Based on the work of this study, musical habits should not be used as a diagnostic tool when trying to identify potential mental disorders in patients, as there is currently no identifiable correlation between said habits and mental well-being.

IX. CHALLENGES

The biggest challenge we faced with this study was undoubtedly the small sample size of our data. The survey used to create the original dataset had only 736 respondents. This was the largest dataset we could find on the subject, but its size is still extremely small for a machine learning task. This, plus necessary preprocessing steps, led to a low-sample and high-feature dataset which is not ideal for model training and testing.

Another challenge faced was determining how to organize the target classes. The original data had ordinal target values ranging from 1-10, which was far too many classes. We initially chose to simplify this into four "severity" categories based on the initial values. This also proved complicated for analyzing our results, as the true positive and false positive rates had to be handled differently than normal for Binary Classes. Altering the data to then be binary, 0 or 1, made it possible to perform a more thorough analysis of the results, though the results ultimately proved to be similar to the Multi-Class results.

Given our results, we were curious to see if other people on Kaggle had done similar work on this topic, especially given little professional research had yet been conducted. After looking through Kaggle, and at a few people's notebooks, it appears that our results are actually on the higher end of other people's work. This goes to show that others have gotten similar results, indicating that our analysis was on par with current findings, but given our accuracy scores were higher, we had made some advancement toward the question at hand.

X. FUTURE WORK

The biggest opportunity to improve on our results in the future, and to better understand if any correlation exists between music and mental health, would be to increase the size of the data. If a larger survey was conducted resulting in more data points, it could be far more effective when it comes to training and testing models, allowing them to better learn the data and avoid overfitting.

We could also consider using a lower threshold for feature to feature correlation when running a Pearson Correlation Matrix for feature selection as this would enable us to drop more features, simplifying the dataset and reducing dimensionality. With a smaller sized dataset, having fewer features can help prevent overfitting - if this truly was the issue we were encountering. However, this may not be the case, as we indicate prior, given our feature importance analysis was low across the board. Regardless, taking a closer look at the data's features and how they could be reduced may help improve the results.

Lastly, using new models that are better built to handle smaller data could also prove to be beneficial. While we used models that ranged in complexity and used models that were meant for smaller datasets, we definitely could have explored other options that may better capture patterns in our data. For example, Gradient Boosting Machines or Support Vector Machines could serve this task well, as well as Logistic Regression which may be effective for our binarized dataset.

REFERENCES

[1] "2014 Nielsen Music U.S. report," Business Wire, https://www.businesswire.com/news/home/20150107006534/en/2014-NIELSEN-MUSIC-U.S.-REPORT.

[2] "2014 Nielsen Music U.S. report," Business Wire, https://www.businesswire.com/news/home/20150107006534/en/2014-NIELSEN-MUSIC-U.S.-REPORT.

[3] T. L. Golden et al., "The use of music in the treatment and management of serious mental illness: A global scoping review of the literature," Frontiers, https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.649840/full.

[4] T. P. A. Truong, B. Applewhite, A. Heiderscheit, and H. Himmerich, "A Systematic Review of Scientific Studies and Case Reports on Music and Obsessive-Compulsive Disorder," International Journal of Environmental Research and Public Health, vol. 18, no. 22, p. 11799, Nov. 2021, doi: https://doi.org/10.3390/ijerph182211799.

[5] Totok Sutojo, et al. "A Machine Learning Approach for Corrosion Small Datasets." Npj, vol. 7, no. 1, 17 Mar. 2023, https://doi.org/10.1038/s41529-023-00336-7.

[6] Olson, Matthew, et al. Modern Neural Networks Generalize on Small Data Sets.