Music and Mind: Examining the Relationship between Music Listening Behaviors and Self-Reported Mental Health

Abstract
Music has been known to improve mood and reduce levels of stress, anxiety, and depression, especially in clinical applications of music therapy. To explore possible factors in music listening patterns that impact people's mental health, this study uses online survey data about music listening and mental health self-evaluations to study whether there is an association between music listening and mental health outcomes while exploring which aspects of listening to music are the most influential towards the impact of music on mental health. A tree model with ensemble methods, binary logistic regression model, and support vector machine models were created to select the optimal model representing the relationship between music listening and respondent mental health. From these models, the support vector classifier model had the most predictive power and the most important variables that affected mental health were listening to music while working, anxiety levels, average BPM, and the respondent's age.

1. Background and Significance

Music has been applied towards improving people's mental health and well-being through music therapy, which Cleveland Clinic describes as listening or interacting with music in a clinical setting towards helping people psychologically, cognitively, or emotionally (Cleveland Clinic). Music therapy has been noted to have various benefits, such as reducing anxiety and depression levels and improving mood. Prior literature on the effects of music on the brain delineate the mechanisms by which music affects the brain, such as activating the auditory cortex, memory regions, and emotional response centers. Through extensively stimulating the brain, music strengthens brain pathways (Budson) and can therefore improve people's mental well-being.

From understanding how listening to music impacts the brain, we are interested in studying different ways music can affect people's mental health based on different genres and types of music along with people's habits when listening to music. Therefore, this study is intended to explore and analyze the factors behind how music can impact people's mental health, especially in respect to listening behaviors and environments.

To address our interest, we proposed the following two research questions:

1. What is the association between different music listening behaviors and people's mental health conditions?
2. What are some influential factors that contribute to this association?

2. Data and Methodology

2.1 Data Description

The dataset, collected by Catherine Rasgaitis from the University of Washington, was sourced from Kaggle and consists of 736 rows and 33 columns. The data was obtained through online survey forms posted on various social media platforms. It includes 3 quantitative variables and 30 categorical variables, with questions related to music listening behaviors such as musical background, listening habits, and genre preference. Participants self-rated their mental health levels on a scale of 0 to 10 for anxiety, depression, insomnia, and OCD. Our response variable of interest is the impact of music on mental health, categorized as either improving, having no effect, or worsening.

2.2 Data Cleaning

The dataset used in this study has 17 missing values from 7 variables, which were removed due to their low percentage (≈2% of the data) and lack of non-random pattern. However, the variable measuring beats per minute (BPM) has 107 missing values, accounting for approximately 15% of the dataset. To address this, regression imputation was used and two outliers in the BPM data were removed. Additionally, the response variable "music effect" had an imbalance issue, with only 17 "worsen" responses compared to 534 "improve" and 165 "no effect" responses. To mitigate this effect, the "no effect" and "worsen" levels were merged into one response level called "no improvement". Then, the dataset was partitioned into training and testing subsets for cross-validation. The final dataset contains 716 observations and 31 variables, with 70% (501 observations) and 30% (215 observations) in the training and testing subsets, respectively.

2.3 Analysis Method

Due to the high number of predictors in the dataset, we employed three modeling methods: (1) Classification tree, namely CART and ensemble methods for their readability, which helps to address the second research question; (2) Binary Logistic Regression for its interpretability on the selected variables' coefficients; and (3) Support Vector Machine for its robustness to multicollinearity issues and efficiency for high-dimensional data. For the tree method, we began by fitting a basic classification tree. After pruning the tree (as shown in Figure 1), we found that the optimal model had only one node, indicating that none of the predictors were utilized in constructing the pruned tree. Therefore, we turned to ensemble methods, specifically bagging and random forest, and evaluated their performances in comparison to the simple tree model by calculating the misclassification rates on the test dataset. AUC values are not calculated because the tree method is not based on probability. Next, we fitted binary logistic regression models. To ensure no multicollinearity issues, VIF screening and Cramer’s V were used, and no multicollinearity issues were detected. Then, stepwise selection with AIC and BIC criteria was employed to determine the optimal logistic models. The log-likelihood of each model was compared, showing that the selected models outperformed the saturated model. Outliers were identified through diagnostic plots and removed; independence was also assumed to hold true for the dataset. After refitting the AIC and BIC models, the misclassification rates were calculated on the optimal discriminant threshold and ROC curves were plotted. The AUC values from the ROC curves were then used to identify the best performing model across all thresholds. Finally, we transformed all categorical variables into dummy variables and implemented the Support Vector Machine method. As SVMs are highly sensitive to parameter selection, we tested a range of cost values and fit Support Vector Classifier and Support Vector Machine (with polynomial and radial kernels) to determine the cost parameter producing the smallest misclassification rate on the test dataset for each model. Additionally, we plotted the ROC curve across all possible discrimination thresholds and computed the area under the curve for each model to identify the optimal model with the best performance across all discrimination thresholds.

3. Results

3.1 Tree

The full tree model consists of 8 terminal nodes, while pruning suggests an insufficient tree model with only 1 node. Comparing the full tree and ensemble methods together (see table 1), the simple tree model and random forest have the lowest misclassification rate on the test dataset (≈0.312).

3.2 Binary Logistic Regression

Based on the sensitivity-specificity plots, the thresholds where sensitivity and specificity are balanced are approximately 0.22 for AIC and 0.225 for BIC (as we do not particularly lean towards any of the responses). Refitting the models and predicting response using the optimal threshold results in the same misclassification rates around 0.39 (as seen in Table 1). However, comparing the AUC values, the refitted AIC binary logistic model outperforms BIC across different thresholds (0.633>0.5693).

3.3 SVM

Through comparing the misclassification rates and the AUC for each model, we found that the SVC and Radial SVM models outperform the Polynomial SVM model by having perfect classification and an AUC of 1. Moreover, since the SVC model involves less parameter tuning, we would prefer SVC than Radial SVM because it produces a more parsimonious model.

4. Conclusion and Other Considerations

4.1 Conclusion

To address our first research question, we identified the SVC model as the optimal model to investigate the association between music listening behaviors and mental health effects. It produces no misclassification, an AUC value of 1, and it is the most parsimonious. However, because SVC models are not interpretable, we do not have any information on important factors contributing to the association, which means we have to look into the outputs of the AIC logistic model and random forest to answer the second research question.

Based on the AIC model, the most important factors that impact one's self-evaluation of their mental health are whether one listens to music while working and one's level of anxiety (significant at the 0.001 level). The coefficients suggest that listening to music while working is associated with a 0.925 decrease in the predicted log-odds of mental health improvement when the other predictors are fixed. Similarly, when one's identified level of anxiety is increased by 1 while all other predictors are held constant, the predicted log-odds of mental health improving decreases by 0.155 (Figure 10). The random forest model, however, identifies "BPM" and "Age" as the most important variables (Figure 3), with a respective 15 and 14 mean decrease in Gini.

4.2 Limitations

Although we merged "no effect" and "worsen" into a single response category, the response variable remains slightly imbalanced, which may lead to biased predictions toward the "improvement" majority response, evident from the initial tree model's confusion matrix showing that 58 instances of "no improvement" were misclassified as "improvement." Additionally, the tree method's overall performance is poor, as shown by the slightly higher misclassification rates of bagging and random forest models compared to the initial tree model. This is due to the lack of pruning in the individual trees used in ensemble methods, which may result in the aggregated trees not outperforming the shorter trees identified by the simple CART algorithm. As a result, the tree method was not considered, limiting the ability to generate a model visualization and identify important variables for addressing the second research question. Another constraint of this research is that the data collected on respondents' mental health conditions are based on their self-assessment, leading to a high level of subjectivity. For instance, the mental health evaluation scale ranging from 0 to 10 may be interpreted differently by different individuals. Furthermore, the initial response variable was divided into three categories, "improve", "no effect", and "worsen", making it difficult to reach a consensus on the precise definition of these outcomes. Possible volunteer bias in the dataset may also limit its representativeness of the general population, as respondents who are music enthusiasts or concerned with their mental health may be overrepresented due to their higher likelihood of participating in a survey on music and mental health.

4.3 Future Improvements

Future studies could address the issue of subjectivity in responses by implementing an index that provides a standardized scale for evaluating mental health. Additionally, to address the mild imbalance that still exists in the response variable, alternative methods to manage imbalanced data should be considered in future explorations of the data, such as choosing other evaluation metrics or using resampling methods.

Appendix

Table 1 - Misclassification rates and AUC of models by methods

Method	Misclassification Rate	AUC
CART	0.3116279	Does not apply
Bagging	0.3162791	Does not apply
Random Forest	0.3116279	Does not apply
AIC Logistic	0.3906977	0.633
BIC Logistic	0.3906977	0.5693
SVC	0	1
Polynomial SVM	0.1767442	0.9978
Radial SVM	0	1

Figure 1

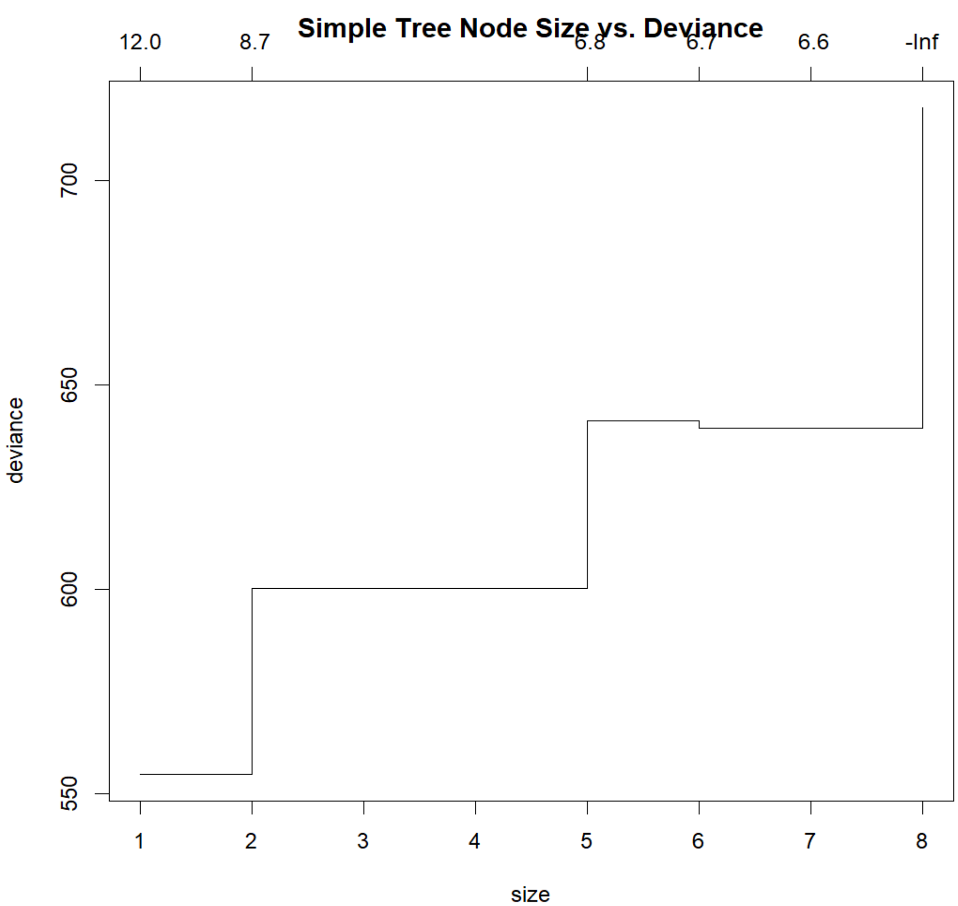


Figure 2

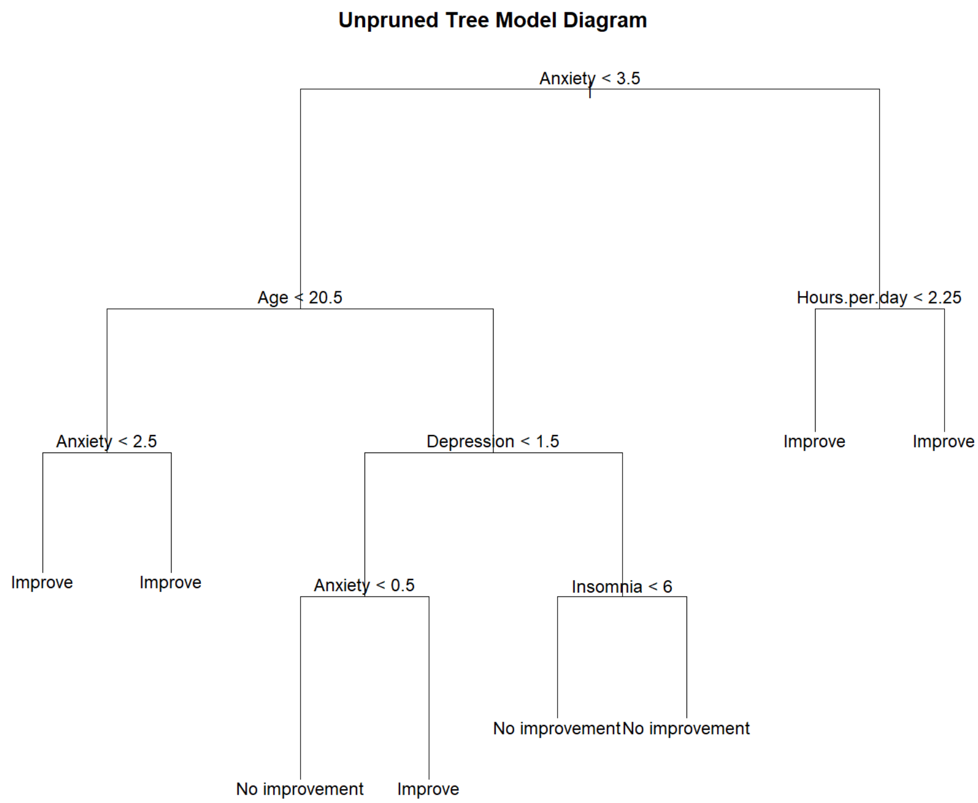


Figure 3

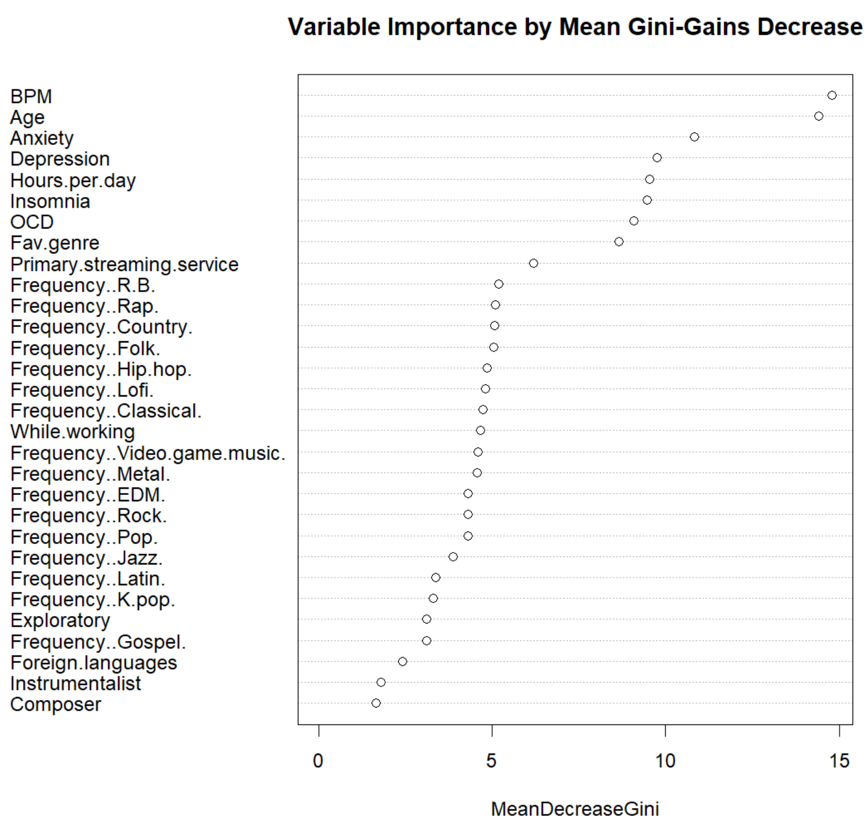
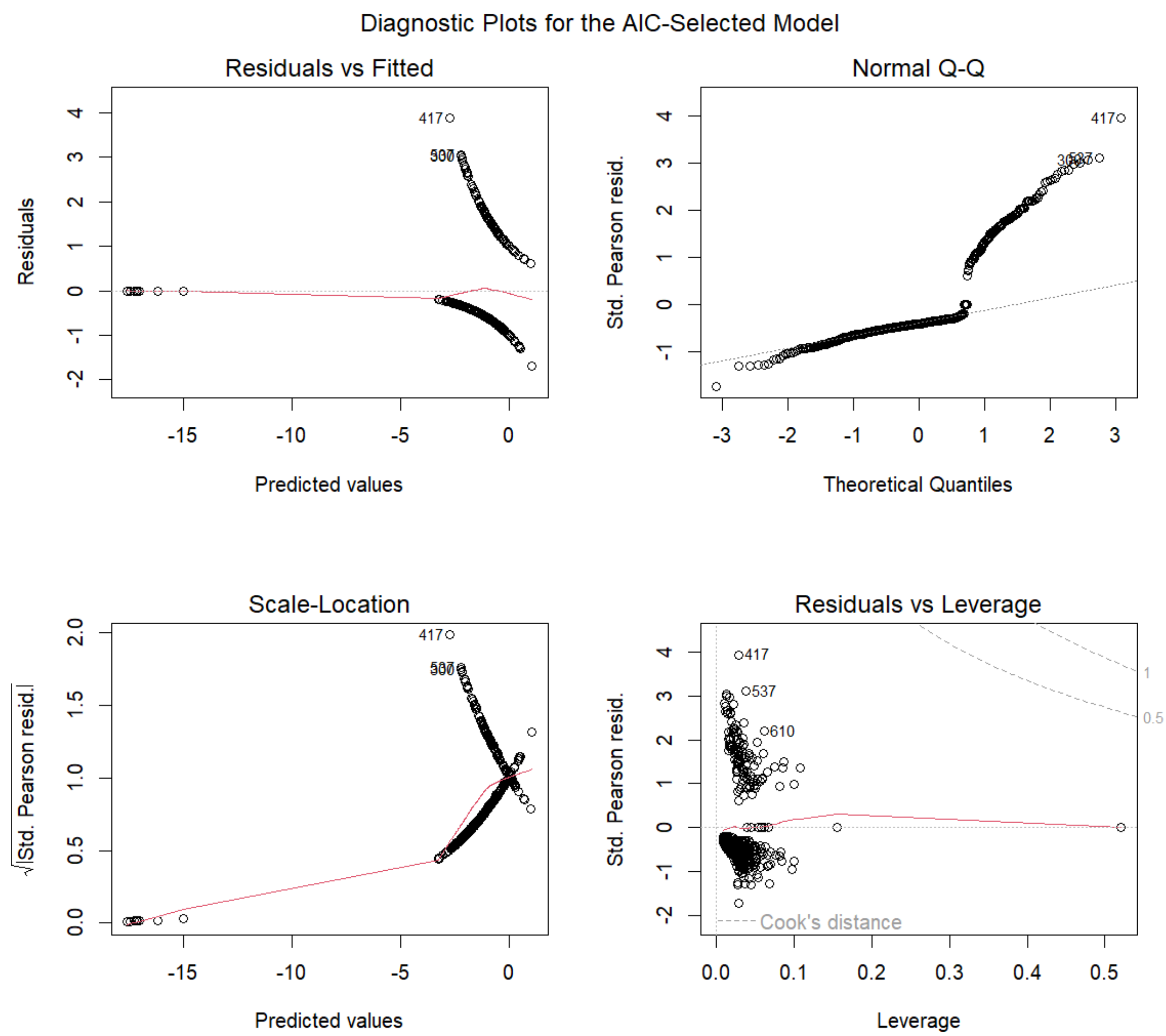


Figure 4



Diagnostic Plots for the AIC-Selected Model
Residuals vs Fitted
Residuals
Predicted values
417
Normal Q-Q
Std. Pearson resid.
Theoretical Quantiles
417
Scale-Location
√|Std. Pearson resid.|
Predicted values
417
Residuals vs Leverage
Std. Pearson resid.
Leverage
417
537
610
Cook's distance
1
0.5

Figure 5

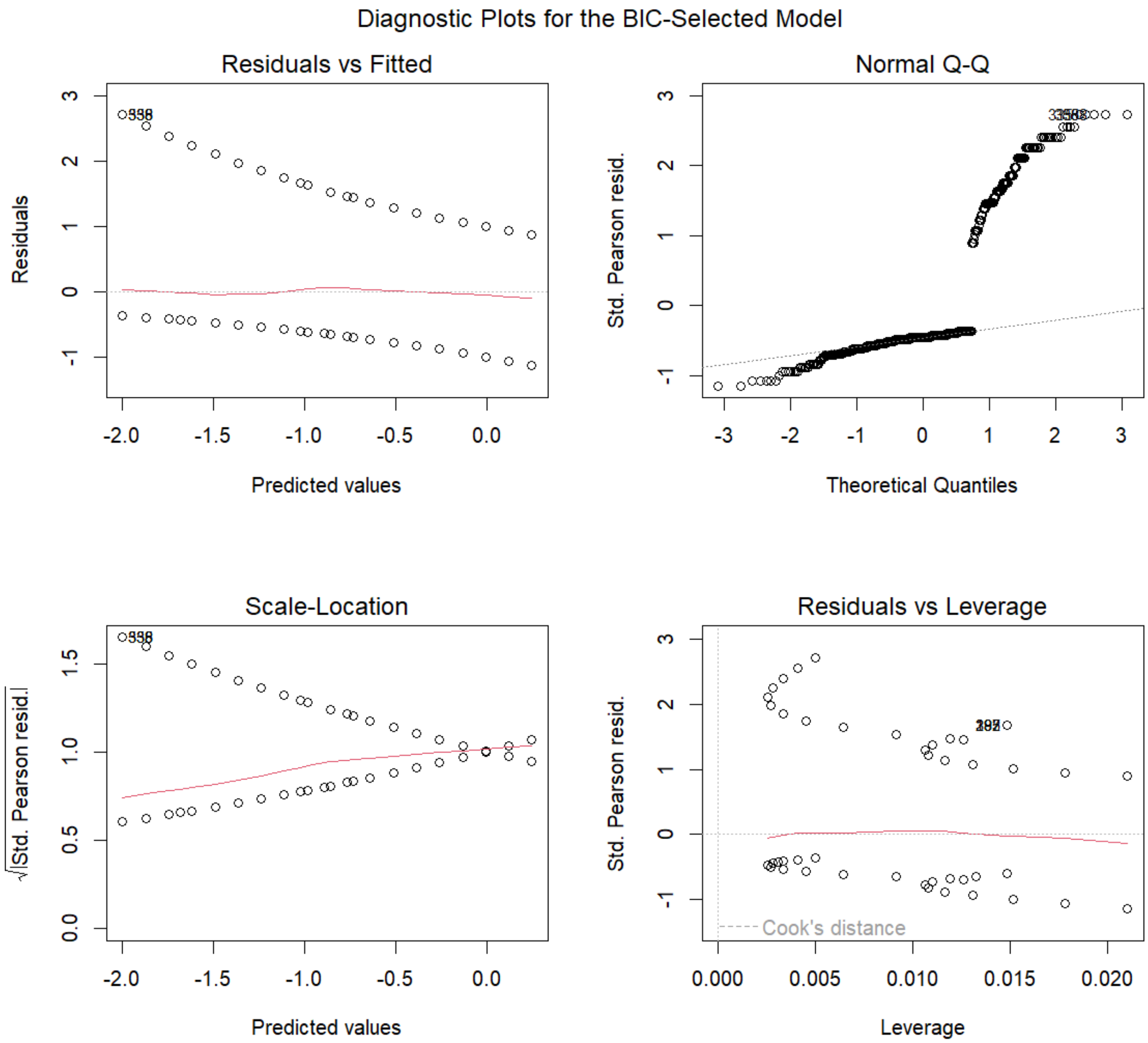

Diagnostic Plots for the BIC-Selected Model
Residuals vs Fitted
Residuals
Predicted values
Normal Q-Q
Std. Pearson resid.
Theoretical Quantiles
Scale-Location
√|Std. Pearson resid.|
Predicted values
Residuals vs Leverage
Std. Pearson resid.
Cook's distance
Leverage

Figure 6

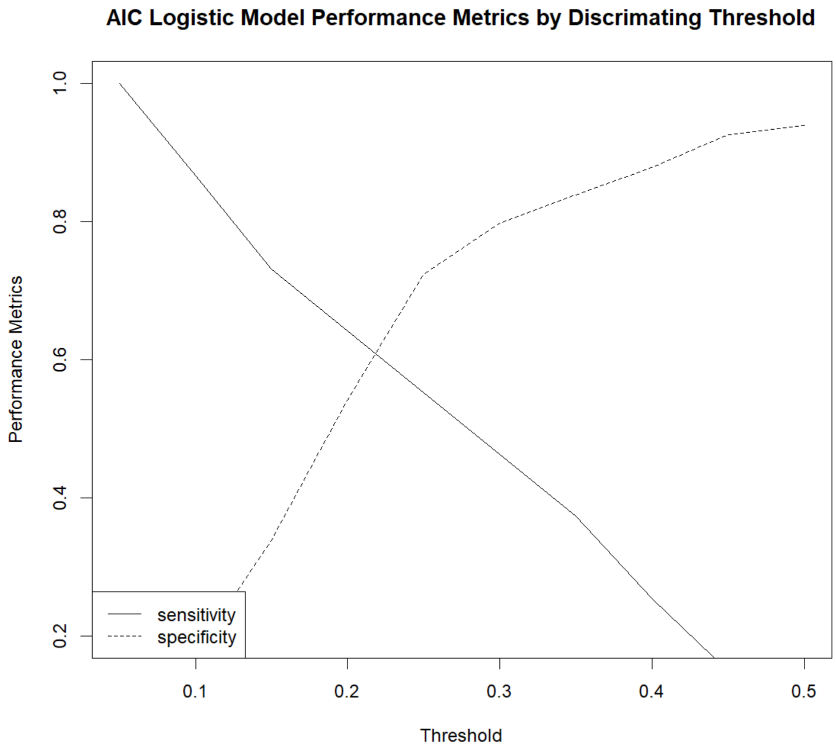


Figure 7

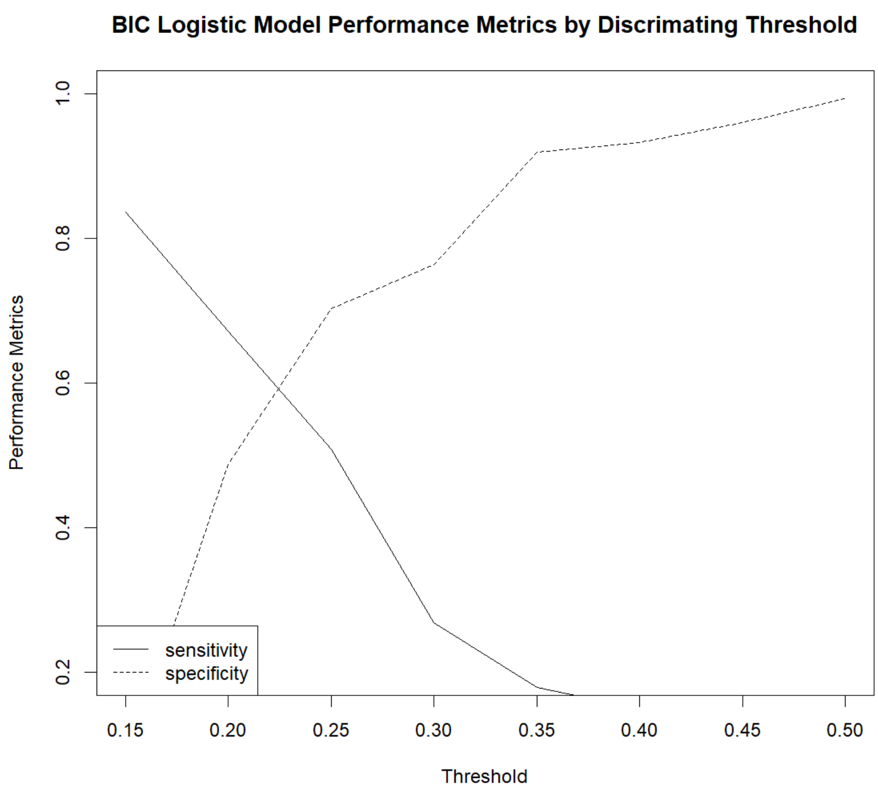


Figure 8

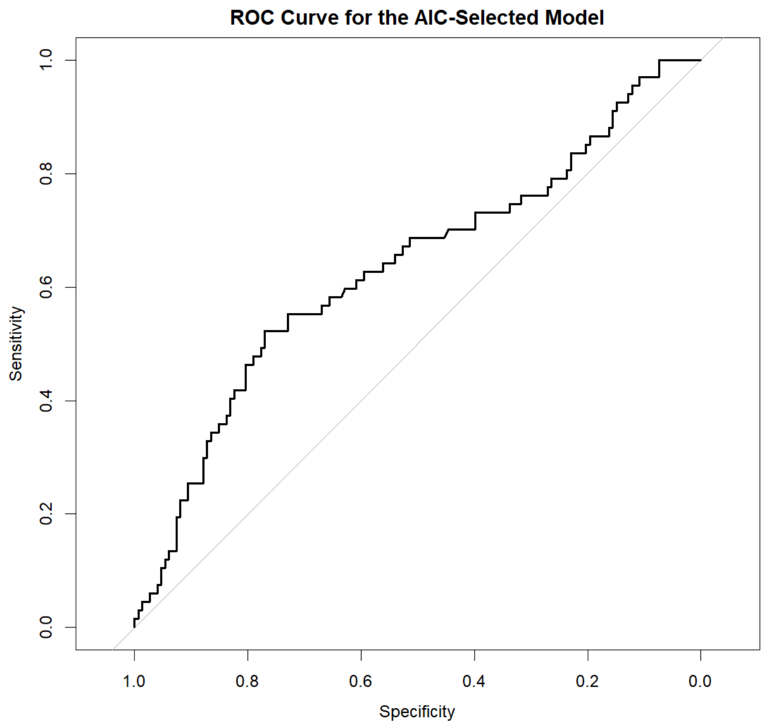


Figure 9

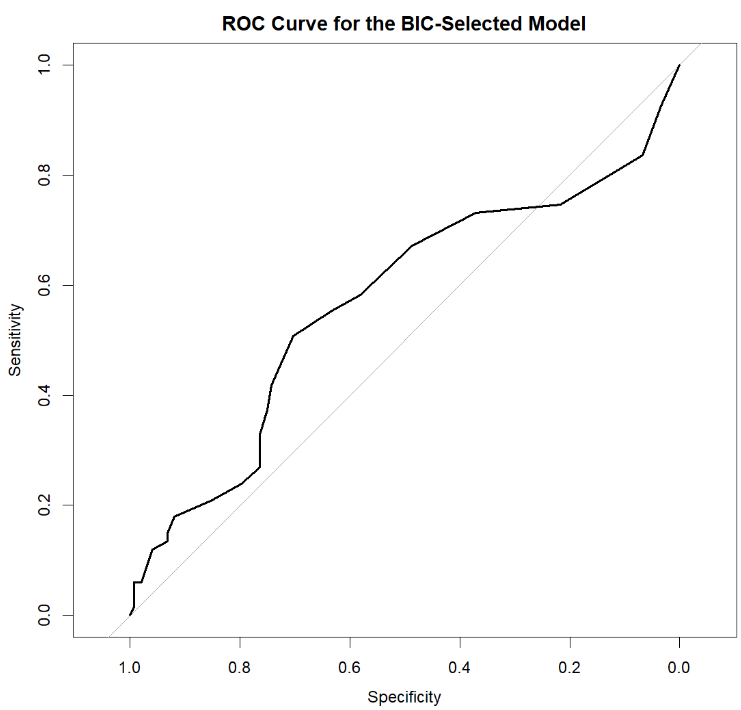


Figure 10 - Output of the refitted AIC model

```
Call:
glm(formula = New.music.effects ~ While.working + Instrumentalist +
    Exploratory + Frequency..Classical. + Frequency..Gospel. +
    Frequency..Jazz. + Anxiety + OCD, family = "binomial", data = music.train[-c(335,
    338, 417, 300, 537, 558, 610), ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6331  -0.7267  -0.5475  -0.2827   2.3409

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                        0.66474    0.40035   1.660 0.096830 .
While.workingYes                  -0.92528    0.26464  -3.496 0.000472 ***
InstrumentalistYes                -0.43704    0.26413  -1.655 0.097999 .
ExploratoryYes                    -0.66253    0.24735  -2.679 0.007395 **
Frequency..Classical.Rarely        0.44015    0.33563   1.311 0.189712
Frequency..Classical.Sometimes     0.72247    0.35090   2.059 0.039502 *
Frequency..Classical.Very frequently  0.84540    0.40898   2.067 0.038724 *
Frequency..Gospel.Rarely          -0.49340    0.32110  -1.537 0.124396
Frequency..Gospel.Sometimes       -0.58037    0.53664  -1.081 0.279486
Frequency..Gospel.Very frequently -16.08438  786.56122  -0.020 0.983685
Frequency..Jazz.Rarely            -0.64273    0.29007  -2.216 0.026706 *
Frequency..Jazz.Sometimes         -0.11239    0.31516  -0.357 0.721389
Frequency..Jazz.Very frequently   -0.41680    0.49962  -0.834 0.404147
Anxiety                           -0.15491    0.04389  -3.530 0.000416 ***
OCD                                0.07761    0.04415   1.758 0.078778 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 11

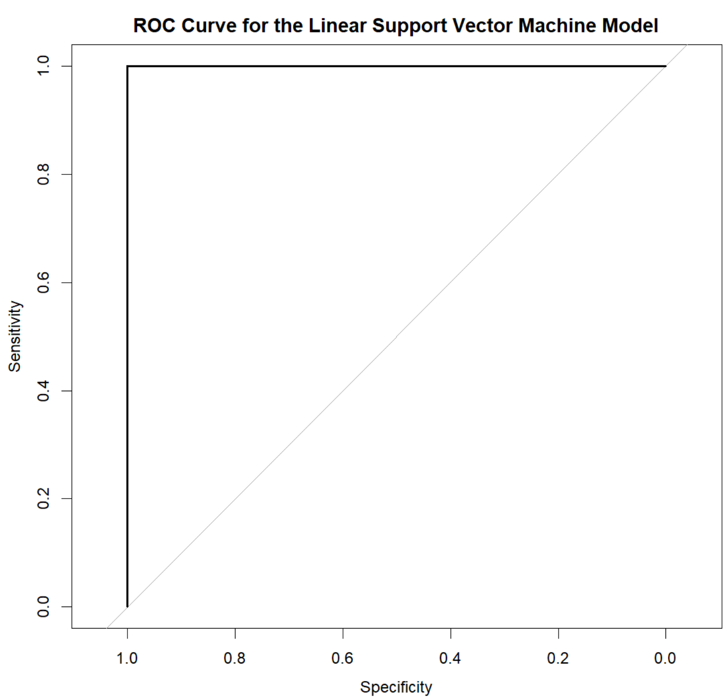


Figure 12

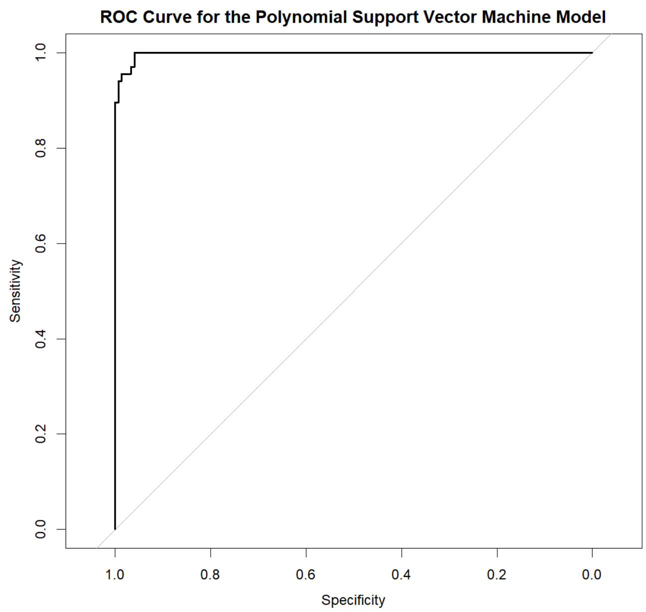
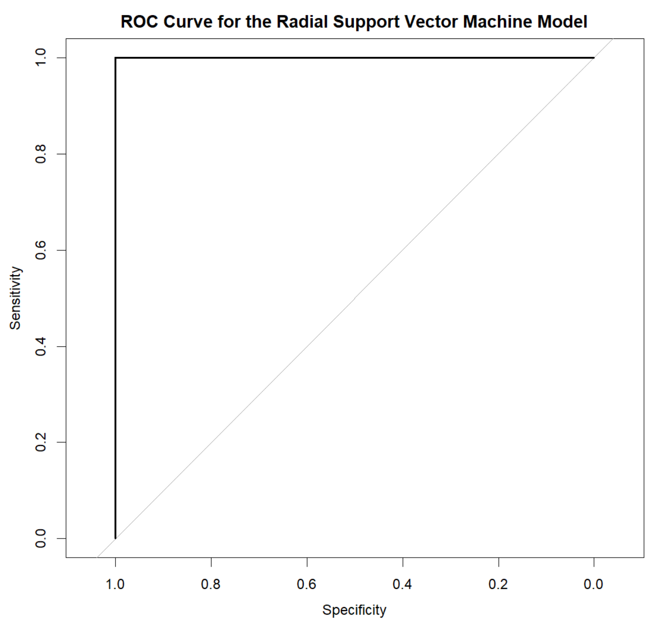


Figure 13



References

Budson, A. E. (2020, October 7). *Why is music good for the brain?* Harvard Health. Retrieved May 5, 2023, from https://www.health.harvard.edu/blog/why-is-music-good-for-the-brain-2020100721062

Music Therapy: What Is It, Types & Treatment. (2020, November 24). Cleveland Clinic. Retrieved May 5, 2023, from https://my.clevelandclinic.org/health/treatments/8817-music-therapy

Rasgaitis, C. (2022, 11). *Music & Mental Health Survey Results*. Kaggle. Retrieved May 5, 2023, from https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results