How music can affect one's mental health based on a set of listening attributes. (Results)

by

Josh Cacayan

Kutztown University

November 25, 2024

Data Science Study using Weka

5.2A - What additional data did you collect during analysis, if any? Include any links or references to the data source.

For this data study, I only used the dataset that was from kaggle.com, which shows how music can impact one's mental health based on listening factors such as music while working, genre type for one's favorite music, and the age of the consumer. No additional data besides the survey on Kaggle was used for this Weka study, which is linked here: Music & Mental Health Survey Results.

5.2B - Intended Goal (Correlation to Music and Mental Health)

The intended goal for this dataset was to check if there was a correlation between whether music affects one's mental health and a set of listening attributes such as their favorite genre, age, the frequency that they listen to music, and their current mental health state (anxiety, depression, etc.). The goal for this dataset was not met, as none of the machine learning models used for the dataset show any correlations between the listening attributes and whether it impacts their mental health or not. The kappa values for all of the classification models used reached a kappa up to only 0.12, which shows that the target value (music effects) can not be determined by some listening attributes. This kappa value was obtained by a Naive Bayes algorithm, which predicted that age is the most important factor of whether music affects one's mental health. The best predictor for the dataset is the age of the consumer and their favorite genre, as there is a 0.04407 correlation between a user's favorite genre and music effects, and a 0.022376 correlation between the user's age and music effect on mental health. However, these predictors, according to the machine learning models, are not very good, as the kappa values for all of them barely have a kappa above .10. The Landis and Koch scale determines the kappa values below 0.20 are slight models, meaning that there isn't any

correlation between the target attribute which is if music impacts one's mental health and other listening factors such as genre, age, and music without working.

5.2.c What machine learning/modeling steps have you taken?

The first step I took when analyzing the dataset was cleaning the dataset I got from kaggle.com found here (Music & Mental Health Survey Results). The unimportant attributes that I removed included the following: Timestamp filling out the survey form, permissions to publish the survey results for each individual, and where a user would primarily listen to their music. Some of the rows in the dataset also needed fixing, as some of the users had no input in regards to whether music would impact their mental health. They would ultimately be removed from the dataset, as it won't contribute to the study. In addition to that, the BPM (beats per minute) column has some null values for some of the rows, and I needed to autofill the bpm with numerical values. Once the data was cleaned, I put the dataset into Weka as an Arff file, to analyze the data with machine learning algorithms. Since the problem I was trying to solve was if any defining factors impacted if music impacts mental health, the models I used were tailored towards classification. The classification models that were used to analyze the dataset were one r, which is good for classifying the biggest factor for a target attribute, a J48 classification tree that highlights the hierarchy of the attributes that impact the target attribute, and a Bayes Net algorithm which classifies how the nontarget attributes in the dataset relate to the target attribute using probabilities.

The results of the machine learning algorithms that I used were not great. The OneR classification model classified the most important attribute as age for impacting one's mental health when listening to music, however, the kappa value for this model is 0.11, which according to the Landis and Koch scale is only slight. The J48 algorithm returns a big decision tree, with the important nodes being the age of the user, if the user uses music to work, and favorite

genre. The age was separated into 2 separate groups, one being people below 14 and the other being people above 14. This model gave a kappa value of 0.08, which is also not great. The best classification model that was used for the dataset was the Bayes Net algorithm, which gave out a kappa value of 0.12.

5.2.D Use SMO, or SMOreg, or multilayer perceptron, or clustering, or at least one other technique not used in assignments 1-3

I also used some algorithms that were not used with the previous assignments, with one being an SMO classification for training the dataset with a support vector machine, and a simple k nearest neighbors algorithm which classifies each instance of the dataset to clusters based on the values of the target attribute. The SMO classification did way worse than the three other classification models, having 200 incorrectly classified instances and a kappa value of around 0.0061. For the K nearest neighbors algorithm, while weka does not display the kappa value for this classification model, the model has around 372 incorrect clustered instances with the clusters being if music improves their mental health or not. This is 50% of the dataset incorrectly classified according to Weka.

5.2.E Revise explaining how could the results of the analysis can be used in a commercial or research setting

While the results of the dataset are not ideal, it shows other researchers that music impacting mental health is not based on certain factors and that for people analyzing music and mental health to be more specific. When I was trying to analyze the dataset to see what certain factors can improve one's mental health when listening to music, I realized that the target attribute for the machine learning model was vague. The target attribute for the machine learning algorithm does not specify how listening to music can improve mental health whether it would help them work on assignments, use it for their emotional states, or help them sleep.

Instead, it would only display if it improves their mental health or has no effect. In addition to this, 74% of the users said that music impacts their mental health while only 25% of people said that music does not improve their mental health/ worsens it. While this may be speculation, it feels like most people who would be interested in filling out the survey would most likely say that music can impact their mental health, and other people who say that music effects don't improve their mental health most likely are not too interested in the study and may have been compelled to fill it out due to extrinsic value. The people who filled out the form may have also listened to a song they hated or loved that day, which could have impacted the result of the dataset.

5.2F Document any other aspect of the project that you feel is important to communicate.

If I were to change anything about analyzing the dataset, I would specify how music would improve or have no effect on their mental health, based on whether or not they would listen to music while working or if listening to music would help them cope with their emotions. In addition to this, I would also reach out to people more based on people who are in the music or psychology industry, as they would be more interested in filling out the survey rather than just the general audience. Overall, music and how it impacts one's mental health unfortunately can not be determined by a specific listening factor according to the results from the machine learning algorithms in Weka.