Correlations Between Aspects of Studying with Music Among College Students

Methodology

In order to reveal the correlations between several aspects of studying with music among college students, I conducted a data analysis on the survey results on seven questions:

(1) What is your major?

(2) What is your GPA?

(3) How often do you listen to music while you study?

(4) How effective do you think listening to music is towards helping you study?

(5) Do you prefer to listen to music significantly more when studying major-related material, slightly more when studying major-related material, slightly more when studying non major-related, significantly more when studying non major related material, or are you indifferent?

(6) What genre of music do you primarily listen to while studying?

(7) What is the level of energy of the music you listen to?

First, I trained a machine learning algorithm called a support vector machine on each of these seven categories, where for each category, I used the data of the remaining six to train the algorithm. The benefit of analyzing the data this way is that this can find nonlinear correlations between more than just two categories. The drawback however is that any machine learning algorithm is usually going to need a lot of data before it can accurately make predictions. After training the support vector machine on each category, I determined the accuracy of its predictions on a cross-validation set. The cross-validation set is a dataset separate from the dataset used to train the algorithm, because if I were to test my algorithm on the same dataset I used to train it, its accuracy would be inflated since it is merely predicting data that it was already given the answer to. After I found the accuracy of its predictions, I had to evaluate the significance of its accuracy by comparing it to the accuracy of simply predicting the most common answer in the dataset. This is important because suppose I made an algorithm that predicts if someone is sick. If 99% of people are healthy, then an algorithm that always says the person is healthy will by 99% accurate, but this is meaningless. So, to account for this, we would need to measure how much *more* accurate the sickness algorithm is than 99%. This is what I did when analyzing my support vector machine.

Once I was done analyzing the support vector machine's ability to predict values for each category, I then evaluated the Pearson r correlation coefficient between every combination of two categories from my survey. The Pearson r correlation coefficient determines the linear correlation between two datasets. In addition, I evaluated the p-value for each correlation coefficient, with the null hypothesis being that the two datasets were not correlated. This just means that I calculated the percent probability that the two categories were not correlated.

Once this was done, I used the results of the correlation coefficients to conduct feature extraction on my support vector machine algorithm. This is important because if the training set has categories that have absolutely nothing to do with the data it is trying to predict, then those unrelated categories can sway the machine learning algorithm from the correct answer. I decided that I would remove categories from a training set that had less than a 50% chance of being linearly correlated to that training set's label (what it is trying to predict). Then I evaluated the performance of the feature extracted version of the support vector machine to determine if the removed features actually contributed to determining that dataset's label.

Potential Biases

The machine learning algorithm needs to be trained on number values, so for most of the categories, preparing the data was straightforward. However, for the categories major and genre, I had to assign a number for each major / genre. This is difficult since there are far too many majors and genres to categorize in only a few categories. But if I categorize them in too many categories, the machine learning algorithm will not have enough data to learn about each category. In addition, in order for the Pearson r correlation coefficient to find a correlation, there must be some order to the values in each category. This meant that I had to decide an order for how to categorize majors and genres, as well as what majors and genres to put in the same category. This is clearly very subjective, and thus a significant area of potential bias. I ended up categorizing them as the following.

By major:

0: Math, Physics

1: Computer Science

2: Engineering, Urban Studies

3: Biology, Ecology, Chemistry, Cognitive Science

4: Art, Music

5: Business, Economics, Communications

6: Sociology, Political Science, Linguistics

7: History, English

8: Ethnic Studies, Gender Studies

By genre:

0: Classical

1: Jazz, R&B

2: Instrumental, Hip Hop

3: Pop

4: Rock, Alt Rock, Electronic Rock

5: Punk, Metal

6: Dubstep, EDM

I made these lists before finalizing the data collected from the surveys, and once all the data was collected, there were no instances of majors in categories 7 or 8. In addition, there were significantly more majors in categories 0 to 3, with the most by far being in 3. This is another area of potential bias since people whose majors are in categories 4 to 8 might tend to answer survey questions differently than people in categories 0 to 3. The survey results for genre however were fairly balanced.

For the sake of clarity and completeness, here is how I ordered the other categories:

(2) 2.0 = 2.0 GPA, to 4.0 = 4.0 GPA

(3) 0 = never listen to music while studying, to 5 = always listen to music while studying

(4) 0 = purely detrimental, to 5 = purely beneficial

(5) 0 = helps major-related significantly more, to 4 = helps non major-related significantly more

(7) 0 = lowest energy music, to 5 = highest energy music

Results

Before I discuss the results, I should point out that I was only able to get 87 survey results—a fairly small sample size. Therefore, any results should be taken with a grain of salt.

For the first analysis of the machine learning algorithm on each category, I got the following results: Category (1) had an accuracy of 45% which was 3% better than predicting the most common major. Category (2) had an accuracy of 25% which was 27% worse than predicting the most common GPA. Category (3) had an accuracy of 40% which was 58% better than predicting the most common answer. Category (4) had an accuracy of 45% which was 68%

better than predicting the most common answer. Category (5) had an accuracy of 65% which was 28% better than predicting the most common answer. Category (6) had an accuracy of 15% which was 33% worse than predicting the most common genre. And category (7) had an accuracy of 45% which was 31% better than predicting the most common answer. This means that we can somewhat accurately predict categories (1), (3), (4), (5), and (7). In other words, each of these things must be based off at least one of the other six qualities in some way. This does not mean categories (2) and (6) are not related to any of the other categories, since it is entirely possible (extremely likely in fact) that my machine learning algorithm does not have enough information to optimally train itself. In order to further evaluate these results, we must look at the results of the Pearson r correlation coefficients (Note that if two things are positively correlated, then if one increases, so does the other. If two things are negatively correlated, then if one increases, the other decreases):

The nine most correlated categories in order from most correlated to least correlated were:

1) (3) and (4) with a 99.99% chance of being positively correlated
2) (6) and (7) with a 99.99% chance of being positively correlated
3) (4) and (7) with a 99.98% chance of being positively correlated
4) (3) and (7) with a 99.98% chance of being positively correlated
5) (4) and (5) with a 98.45% chance of being negatively correlated
6) (2) and (3) with a 95.14% chance of being negatively correlated
7) (4) and (6) with a 94.86% chance of being positively correlated
8) (3) and (6) with a 91.87% chance of being positively correlated
9) (3) and (5) with an 88.45% chance of being negatively correlated

The three least correlated categories in order from least correlated to most correlated were:

1) (1) and (5) with a 94.93% chance of being uncorrelated
2) (1) and (6) with a 94.24% chance of being uncorrelated
3) (5) and (6) with a 75.07% chance of being uncorrelated

With forty-two different ways to pair the seven categories, most correlation coefficients were inconclusive. However, some of the correlations that were found were very surprising. In particular, I did not expect “GPA” and “how often a person listens to music while studying” to be negatively correlated (correlation 6). In fact, I fully expected there to be a positive correlation between the two. One possible interpretation is that people who need music to help them study probably do not enjoy studying in the first place, which is why they use music to help them focus. And since they do not enjoy studying anyways, they tend to have worse GPAs. Another possible interpretation is that music distracts people from studying, making them less efficient at studying, thus lowering their GPAs. However, it is impossible to tell based solely off these survey results.

Another interesting result is the correlation between “the level of energy in music a person prefers” and “how often a person listens to music while studying” (correlation 4). This correlation is particularly strong, which I did not expect at all. As someone who tends to prefer slightly lower energy music, it comes as a surprise that people who prefer higher energy music tend to listen to music more often while studying.

An interesting correlation to note is there was a 74.27% chance that “the level of energy in music a person prefers” and “GPA” are negatively correlated, but this is actually expected as a result of the last two correlations mentioned. People who prefer lower energy music tend to listen to music less often while studying, and people who listen to music less often while studying tend to have better GPAs.

Another example of a correlation that arose through other correlations is the correlation between "how beneficial a person thinks studying with music is" and "what genre of music a person prefers" (correlation 7). The genre a person prefers is strongly correlated with how much energy a person prefers in their music, and people who prefer higher energy music tend to study with music more often, and people who study with music more often tend to think it is beneficial to study with music.

Some of the results are expected, but interesting nonetheless. For example, "how often a person listens to music when they study" is negatively correlated to "whether a person prefers to listen to music as they study major-related or non major-related material" (correlation 9), which makes sense since you would expect people to study material for their major more often, so if a person prefers to listen to music when they study major-related material, they will likely say they listen to music more often.

After analyzing the correlation coefficients, I removed from the machine learning algorithms the features that had more than a 50% chance of not being linearly correlated to what they were predicting. For category (1), this was (5) and (7). For category (2), this was category (5). For categories (3) and (4), nothing was removed. For category (5), this was (1), (2), (5), (6), and (7). For category (6), this was (1) and (5). For category (7), this was (5). So if the feature selected version of the machine learning algorithm does better, that means there exist nonlinear correlations between the category being predicted and at least one of the removed categories. On the feature selected algorithms, category (1) had an accuracy of 40% which was 8% worse than predicting the most common major. Category (2) had an accuracy of 25% which was 27% worse than predicting the most common GPA. For categories (3) and (4) nothing changed since no categories were removed. Category (5) had an accuracy of 40% which was 21% worse than predicting the most common answer. Category (6) had an accuracy of 25% which was 12% better than predicting the most common genre. And category (7) had an accuracy of 40% which was 17% better than predicting the most common answer. So as it turns out, categories (1), (5), and (7) did worse once the features that likely had no linear correlation were removed. At first glance, this seems to imply that at least one of the removed features has some correlation to the category being predicted. However, it is very likely that the difference in performance is due to random chance, since the dataset I had available to me was not very large, and the difference in performance between the original machine learning algorithm and the feature extracted one is very small. So even if there were correlations outside that found by the correlation coefficients, they would not be very significant ones. As a result, it is fair to judge the correlations between categories using only the correlation coefficients, given the amount of data we have to work with.