

How music impacts mental health from a set of listening attributes.

Josh Cacayan



Intended Goals for this study

- I believed that the study that was used for the dataset was interesting, as one who listens to lots of music and likes psychology.
- Does Music have any effect on mental health and can it be predicted? **What factors contribute?**
- Data analysis has been done, however no Machine Learning Algorithms used



Dataset used

- Music and Mental Health Survey Results
- Found on kaggle.com!!
- [Music & Mental Health Survey Results](#)



kaggle™

Cleaning the Data

- Used excel to help clean the data
 - Removed rows where 'music effects' are not present
 - Removed Columns that were not useful (Timestamp, Primary Streaming, Permissions)
 - Autofilled a few null values such as some bpms, and age
- Turned the excel file into an arff file



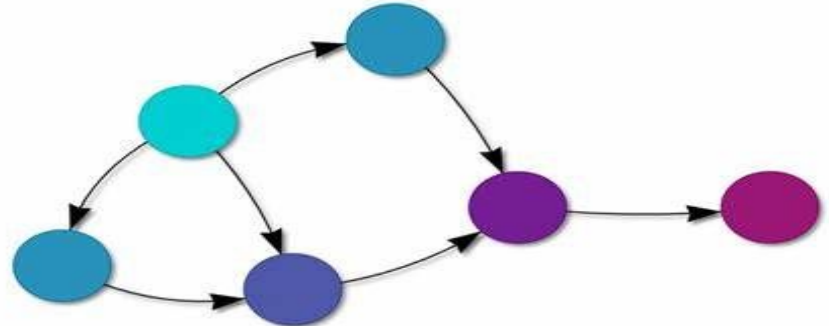
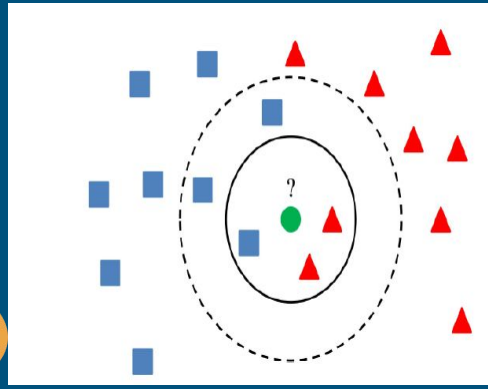
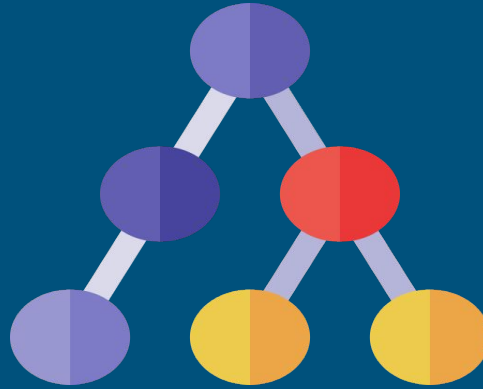
Machine Learning Algorithms

Used in previous assignments

- Bayes Net Algorithm
- J48 Decision Tree
- One R Rule

Not used in previous assignments

- SMO Function
- Simple KNN Means

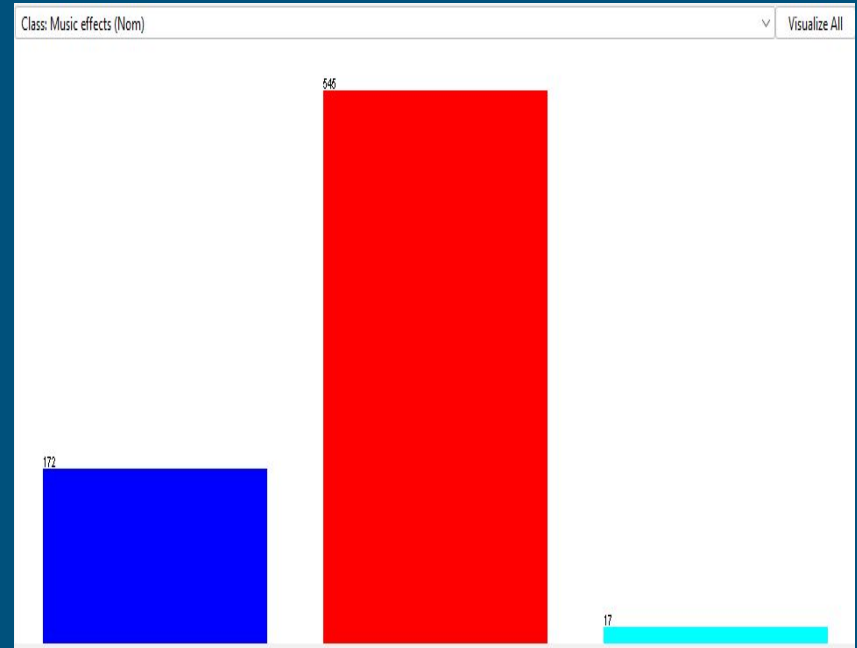


Attributes Used (Target)

Target Attribute: Music Effects

(Does music improve mental health)

- Improve (545 instances)
- No Effect (172 instances)
- Worsen (17 instances)



Attributes Used (Non-Target)

- Age
- Hours Per Day
- While Working
- Composer and Instrumentalist
- Favorite Genre
- BPM (Beats Per minute)
- Frequency (Genres)
- Mental Health
 - OCD
 - Depression
 - Insomnia
 - Anxiety

No.	Name
1	<input type="checkbox"/> Age
2	<input type="checkbox"/> Hours per day
3	<input type="checkbox"/> While working
4	<input type="checkbox"/> Instrumentalist
5	<input type="checkbox"/> Composer
6	<input type="checkbox"/> Fav genre
7	<input type="checkbox"/> Exploratory
8	<input type="checkbox"/> Foreign languages
9	<input type="checkbox"/> BPM
10	<input type="checkbox"/> Frequency [Classical]
11	<input type="checkbox"/> Frequency [Country]
12	<input type="checkbox"/> Frequency [EDM]
13	<input type="checkbox"/> Frequency [Folk]
14	<input type="checkbox"/> Frequency [Gospel]
15	<input type="checkbox"/> Frequency [Hip hop]
16	<input type="checkbox"/> Frequency [Jazz]
17	<input type="checkbox"/> Frequency [K pop]
18	<input type="checkbox"/> Frequency [Latin]
19	<input type="checkbox"/> Frequency [Lofi]
20	<input type="checkbox"/> Frequency [Metal]
21	<input type="checkbox"/> Frequency [Pop]
22	<input type="checkbox"/> Frequency [R&B]
23	<input type="checkbox"/> Frequency [Rap]
24	<input type="checkbox"/> Frequency [Rock]
25	<input type="checkbox"/> Frequency [Video game music]
26	<input type="checkbox"/> Anxiety
27	<input type="checkbox"/> Depression
28	<input type="checkbox"/> Insomnia
29	<input type="checkbox"/> OCD
30	<input type="checkbox"/> Music effects

Findings weren't great..



One-R Algorithm

```
Age:
  < 25.5 -> Improve
  < 26.5 -> No Effect
  < 33.5 -> Improve
  < 35.5 -> No Effect
  < 50.5 -> Improve
  < 56.5 -> No Effect
  < 84.5 -> Improve
  >= 84.5 -> No Effect
(556/734 instances correct)

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      551           75.0681 %
Incorrectly Classified Instances    183           24.9319 %
Kappa statistic                     0.1194
Mean absolute error                  0.1662
Root mean squared error              0.4077
Relative absolute error              63.2016 %
Root relative squared error         112.6033 %
Total Number of Instances          734
```

- Determines the most important attribute: **Age**
- 75% correct instances, 24% incorrect
- Kappa Value: 0.11 (Slight)
- This is considered a slight (bad) Model in Landis and Koch Terms

J48 Algorithm

```
Age <= 14
| Depression <= 9: Improve (27.0/2.0)
| Depression > 9: Worsen (2.0)
Age > 14
| While working = Yes
| | Depression <= 0
| | | Frequency [Folk] = Never
| | | | Frequency [Latin] = Very frequently: No Effect (3.0/1.0)
| | | | Frequency [Latin] = Sometimes
| | | | | Insomnia <= 5: Improve (4.0)
| | | | | Insomnia > 5: No Effect (2.0)
| | | | Frequency [Latin] = Never
| | | | | OCD <= 2: No Effect (14.78/4.0)
| | | | | OCD > 2: Improve (3.0)
| | | | Frequency [Latin] = Rarely: Improve (4.0)
| | | Frequency [Folk] = Rarely
| | | | Age <= 20: No Effect (2.0)
| | | | Age > 20: Improve (12.0/1.0)
| | | Frequency [Folk] = Sometimes: Improve (7.0/1.0)
| | | Frequency [Folk] = Very frequently: No Effect (4.0)
| | Depression > 0: Improve (496.78/101.0)
| While working = No
| | Fav genre = Latin: Improve (0.0)
| | Fav genre = Rock
| | | Instrumentalist = Yes: Improve (10.0/2.0)
| | | Instrumentalist = No
| | | | Frequency [Latin] = Very frequently: No Effect (0.0)
| | | | Frequency [Latin] = Sometimes: Improve (2.0/1.0)
| | | | Frequency [Latin] = Never
| | | | | Frequency [K pop] = Very frequently: No Effect (1.0)
| | | | | Frequency [K pop] = Rarely: Improve (4.0)
| | | | | Frequency [K pop] = Sometimes: No Effect (1.0)
| | | | | Frequency [K pop] = Never
| | | | | Frequency [Pop] = Very frequently: Improve (4.0)
| | | | | Frequency [Pop] = Sometimes
| | | | | | Foreign languages = Yes: Improve (2.0)
| | | | | | Foreign languages = No
| | | | | | | Frequency [Lofi] = Rarely: Improve (3.0/1.0)
| | | | | | | Frequency [Lofi] = Sometimes: Improve (1.0)
| | | | | | | Frequency [Lofi] = Very frequently: No Effect (0.0)
| | | | | | | Frequency [Lofi] = Never: No Effect (6.0)
| | | | | Frequency [Pop] = Rarely: No Effect (4.22)
| | | | | Frequency [Pop] = Never: No Effect (0.0)
| | | Frequency [Latin] = Rarely
| | | | Age <= 33: No Effect (4.0)
| | | | Age > 33: Improve (2.0)
| Fav genre = Video game music: No Effect (8.0/2.0)
| Fav genre = Jazz: No Effect (3.0)
```

```
Number of Leaves : 71
Size of the tree : 99
```

Time taken to build model: 0.01 seconds

```
=== Stratified cross-validation ===
=== Summary ===
```

Correctly Classified Instances	532	72.4796 %
Incorrectly Classified Instances	202	27.5204 %
Kappa statistic	0.0848	
Mean absolute error	0.2451	
Root mean squared error	0.381	
Relative absolute error	93.1916 %	
Root relative squared error	105.2221 %	
Total Number of Instances	734	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	0.134	0.068	0.377	0.134	0.197	0.101
	0.934	0.862	0.757	0.934	0.836	0.112
	0.000	0.001	0.000	0.000	0.000	-0.006
Weighted Avg.	0.725	0.656	0.651	0.725	0.667	0.107

```
=== Confusion Matrix ===
```

a	b	c	<-- classified as
23	149	0	a = No Effect
35	509	1	b = Improve
3	14	0	c = Worsen

SMO (Support Vector Machine)

```
+ 0.0255 * (normalized) Frequency [R&B]=Never
+ 0.0093 * (normalized) Frequency [R&B]=Very frequently
+ 0.458 * (normalized) Frequency [R&B]=Rarely
+ 0.1407 * (normalized) Frequency [Rap]=Very frequently
+ -0.0841 * (normalized) Frequency [Rap]=Rarely
+ -0.208 * (normalized) Frequency [Rap]=Never
+ 0.1514 * (normalized) Frequency [Rap]=Sometimes
+ 0.3236 * (normalized) Frequency [Rock]=Never
+ -0.2757 * (normalized) Frequency [Rock]=Very frequently
+ -0.1992 * (normalized) Frequency [Rock]=Rarely
+ 0.1512 * (normalized) Frequency [Rock]=Sometimes
+ -0.3041 * (normalized) Frequency [Video game music]=Sometimes
+ -0.1907 * (normalized) Frequency [Video game music]=Rarely
+ 0.3141 * (normalized) Frequency [Video game music]=Very frequently
+ 0.1807 * (normalized) Frequency [Video game music]=Never
+ -0.1258 * (normalized) Anxiety
+ 1.4042 * (normalized) Depression
+ -0.0413 * (normalized) Insomnia
+ 0.1847 * (normalized) OCD
- 2.4794
```

Number of kernel evaluations: 51535 (90.90% cached)

Time taken to build model: 0.63 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	534	72.752 %
Incorrectly Classified Instances	200	27.248 %
Kappa statistic	0.0061	
Mean absolute error	0.2925	
Root mean squared error	0.3792	
Relative absolute error	111.2072 %	
Root relative squared error	104.7398 %	
Total Number of Instances	734	

- Classifies the target attribute (Music Effects) using a hyperplane to separate instances to classes based on distance
- 72% correct instances, 27% incorrect
- Kappa Value: 0.0061 (Slight)
- This is considered a slight (bad) Model in Landis and Koch Terms

Bayes Net Algorithm

```
Age(2): Music effects
Hours per day(1): Music effects
While working(2): Music effects
Instrumentalist(2): Music effects
Composer(2): Music effects
Fav genre(16): Music effects
Exploratory(2): Music effects
Foreign languages(2): Music effects
BPM(1): Music effects
Frequency [Classical](4): Music effects
Frequency [Country](4): Music effects
Frequency [EDM](4): Music effects
Frequency [Folk](4): Music effects
Frequency [Gospel](4): Music effects
Frequency [Hip hop](4): Music effects
Frequency [Jazz](4): Music effects
Frequency [K pop](4): Music effects
Frequency [Latin](4): Music effects
Frequency [Lofi](4): Music effects
Frequency [Metal](4): Music effects
Frequency [Pop](4): Music effects
Frequency [R&B](4): Music effects
Frequency [Rap](4): Music effects
Frequency [Rock](4): Music effects
Frequency [Video game music](4): Music effects
Anxiety(1): Music effects
Depression(1): Music effects
Insomnia(1): Music effects
OCD(1): Music effects
Music effects(3):
LogScore Bayes: -19418.772482334323
LogScore BDeu: -19935.480485904693
LogScore MDL: -19916.75241005569
LogScore ENTROPY: -19227.20821656547
LogScore AIC: -19436.20821656547

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      505           68.8011 %
Incorrectly Classified Instances    229           31.1989 %
Kappa statistic                    0.1267
Mean absolute error                 0.2506
Root mean squared error             0.3854
Relative absolute error             95.2932 %
Root relative squared error         105.4522 %
```

- Algorithm that classifies how nontarget attributes relate to the target attribute by finding probabilities with bayes theorem.
- 68% correct instances, 31% incorrect instances
- Kappa: 0.12% (Slight)
- This is considered a slight (bad) Model in Landis and Koch Terms

Simple K Means Algorithm

```
Time taken to build model (full training data) : 0.05 seconds
```

```
=== Model and evaluation on training set ===
```

```
Clustered Instances
```

```
0      221 ( 30%)
1      242 ( 33%)
2      271 ( 37%)
```

```
Class attribute: Music effects
```

```
Classes to Clusters:
```

```
  0   1   2  <-- assigned to cluster
49  45  78 | No Effect
168 192 185 | Improve
  4   5   8 | Worsen
```

```
Cluster 0 <-- Worsen
```

```
Cluster 1 <-- Improve
```

```
Cluster 2 <-- No Effect
```

```
Incorrectly clustered instances :      460.0      62.6703 %
```

- Categorized all instances to 3 classes:
 - Cluster 0: Improve Mental Health
 - Cluster 1: No Effect on Mental Health
 - Cluster 2: Worsen Mental Health
- Clustered Instances
 - Improve Mental Health: 221 instances
 - No effect on Mental Health: 242 instances
 - Worsen Mental Health: 271 instances
- Incorrect clustered instances: 62%!!! (Worse model yet)

What Went Wrong?



Possible Problems of the Dataset

- Dataset was unbalanced, 74% of users say it would improve, 25% has no effect/worsens mental health
- Vague in determining how it impacts a person's mental health
- Dataset was made as a survey, with the statistician having no control of external factors



What can be taken away?

- Most people that listen to music most likely will improve their mental health!
- There are no specific attributes on what determines if music impacts mental health or not.
- A dataset which was formed with a survey is not the greatest for classification analysis (Kappa values were low)
- Survey should be specific on what mental health improvements entails
 - If it helps cope with their emotions
 - If the user listens to music while working does it help?



Any Questions?????

