

Available Data:

Trial 1: data_3_8_16.txt

Trial 2: data_3_1_16.txt

Trial 3: data_2_16_16.txt

Trial 4: data_2_19_16.txt

Trial 5: data_2_8_16.txt

Users:

1246241522	1257662670
connor	Corne
hunter5474	jungleprince
mlhopp	naomimusgrave
sana	svetlang

Musical Features: danceability, energy, liveness, loudness, speechiness, tempo, valence, instrumentalness, acousticness

Hypothesis:

Each data set consist of the following: a user, a user's music playlist, and Spotify's API audio features (or description) for each song. Our goal is to implement the unsupervised model Artificial Neural Network with Logistic Regression machine learning model, in order to classify the musical taste of an individual user, based on the musical features of user's playlist. For each song, its features are as follows: danceability, energy, liveness, loudness, speechiness, tempo, valence, instrumentalness, acousticness. In an effort to further our understanding of an individual's music preference, we will reduce the number of feature that our model will run, in hopes of narrowing down and determining which features a particular user enjoys more.

Conclusion:

- The higher the number of features represented, the more general the classification of users becomes. When certain users are classified with high accuracy under this conditions, we deduct these are users of broad musical preference.
- Examples: user 1246241522 in several occasions have shown signs of overfitting, this is because despite running the model with several different subsets of feature data, user 1246241522 was identified consistently with higher accuracy that reach nearly 90% to 100%. With this in mind, the features selected in the second trial for this user were strong evidence towards our hypothesis, and demonstrated across 80% of the data.
- Our hypothesis is correct to a 30%-40% accurate depending on the features represented per trial. Like mentioned before the reduction turned to a higher amount of features for classifying people. A more robust method for reducing the feature space could have been used in order to observe the consistency of the hypothesis. In combination with

normalization of data and distributed along a wider variety of users musical playlist, the classification result would have been improved. From our results we can observe that in the range where we do not obtain perfect classification through selected features, with 50%-70% certainty we obtained the users in the top three most accurate result.

Experiment:

Trial Part 1: Run ann with powerset [true] with all features

Trial Part 2: Run ann with powerset [false] with all features

Trial Part 3: Run ann with powerset [false] with best feature set from part 1

Trial Part 4 sections a through d: Run ann with powerset [false] with best feature set from part 1 on the remaining 4 other data sets.

Trail 1 Part 1:

```
python run.py ann 1 true data/data_3_8_16.txt > ann1_1.txt
```

Results:

Accuracy: 42 correct gives 0.3 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy', 'liveness', 'loudness', 'speechiness', 'tempo', 'valence', 'instrumentalness', 'acousticness']

Identified 7 out of 7 labels

naomimusgrave	11	of 18	(0.611111)
hunter5474	3	of 22	(0.136364)
jungleprince	13	of 24	(0.541667)
connor	4	of 20	(0.200000)
mlhopp	2	of 17	(0.117647)
1246241522	4	of 20	(0.200000)
sana	5	of 19	(0.263158)

Trail 1 Part 2:

```
python run.py ann 1 false data/data_3_8_16.txt > ann1_2.txt
```

Results:

Accuracy: 38 correct gives 0.271428571429 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy', 'liveness', 'loudness', 'speechiness', 'tempo', 'valence', 'instrumentalness', 'acousticness']

Identified 7 out of 7 labels

naomimusgrave	4	of 17	(0.235294)
hunter5474	4	of 18	(0.222222)
jungleprince	12	of 23	(0.521739)
connor	1	of 25	(0.040000)
mlhopp	2	of 18	(0.111111)
1246241522	10	of 18	(0.555556)
sana	5	of 21	(0.238095)

Trail 1 Part 3:

```
python run.py ann 1 false data/data_3_8_16.txt > ann1_3.txt
```

Results:

Accuracy: 39 correct gives 0.278571428571
 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy',
 'liveness', 'loudness', 'speechiness', 'tempo', 'valence',
 'instrumentalness', 'acousticness']

Identified 6 out of 7 labels

naomimusgrave	3	of 20	(0.150000)
hunter5474	4	of 20	(0.200000)
jungleprince	15	of 26	(0.576923)
connor	7	of 18	(0.388889)
1246241522	7	of 19	(0.368421)
sana	3	of 17	(0.176471)

Trail 1 Part 4a:

```
python run.py ann 1 false data/data_2_16_16.txt > ann1_4a.txt
```

Results:

Accuracy: 32 correct gives 0.213333333333
 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy',
 'liveness', 'loudness', 'speechiness', 'tempo', 'valence',
 'instrumentalness', 'acousticness']

Identified 2 out of 7 labels

jungleprince	6	of 23	(0.260870)
1246241522	26	of 26	(1.000000)

Trail 1 Part 4b:

```
python run.py ann 1 false data/data_2_19_16.txt > ann1_4b.txt
```

Results:

Accuracy: 43 correct gives 0.286666666667
compared to guessing 0.142857142857
Best feature set: ['danceability', 'energy',
'liveness', 'loudness', 'speechiness', 'tempo', 'valence',
'instrumentalness', 'acousticness']
Identified 3 out of 7 labels

jungleprince	4	of 22	(0.181818)
1246241522	37	of 40	(0.925000)
naomimusgrave	2	of 30	(0.066667)

Trail 1 Part 4c:

```
python run.py ann 1 false data/data_2_8_16.txt > ann1_4c.txt
```

Results:

Accuracy: 28 correct gives 0.466666666667
compared to guessing 0.333333333333
Best feature set: ['danceability', 'energy',
'liveness', 'loudness', 'speechiness', 'tempo', 'valence',
'instrumentalness', 'acousticness']
Identified 3 out of 3 labels

mlhopp	7	of 18	(0.388889)
naomimusgrave	12	of 25	(0.480000)
connor	9	of 17	(0.529412)

Trail 1 Part 4d:

```
python run.py ann 1 false data/data_3_1_16.txt > ann1_4d.txt
```

Results:

Accuracy: 58 correct gives 0.256637168142
compared to guessing 0.0909090909091
Best feature set: ['danceability', 'energy',
'liveness', 'loudness', 'speechiness', 'tempo', 'valence',
'instrumentalness', 'acousticness']
Identified 4 out of 11 labels

jungleprince	17	of 36	(0.472222)
corne	1	of 16	(0.062500)
1246241522	39	of 41	(0.951220)
naomimusgrave	1	of 27	(0.037037)

Trail 2 Part 1:

```
python run.py ann 1 true data/data_3_1_16.txt > ann2.txt
```

Results:

Accuracy: 55 correct gives 0.243362831858
compared to guessing 0.0909090909091

Best feature set: ['danceability', 'energy',
'liveness', 'instrumentalness', 'acousticness']

Identified 2 out of 11 labels

jungleprince	22	of 34	(0.647059)
1246241522	33	of 33	(1.000000)

Trail 2 Part 2:

```
python run.py ann 1 false data/data_3_1_16.txt > ann2_2.txt
```

Results:

Accuracy: 51 correct gives 0.225663716814
compared to guessing 0.0909090909091

Best feature set: ['danceability', 'energy',
'liveness', 'loudness', 'speechiness', 'tempo', 'valence',
'instrumentalness', 'acousticness']

Identified 2 out of 11 labels

jungleprince	17	of 24	(0.708333)
1246241522	34	of 36	(0.944444)

Trail 2 Part 3:

```
python run.py ann 1 false data/data_3_1_16.txt danceability  
energy liveness instrumentalness acousticness > ann2_3.txt
```

Results:

Accuracy: 42 correct gives 0.185840707965
compared to guessing 0.0909090909091

Best feature set: ['danceability', 'energy',
'liveness', 'instrumentalness', 'acousticness']

Identified 2 out of 11 labels

jungleprince	17	of 29	(0.586207)
1246241522	25	of 27	(0.925926)

Trail 2 Part 4a:

```
python run.py ann 1 false data/data_3_8_16.txt danceability  
energy liveness instrumentalness acousticness > ann2_4a.txt
```

Results:

Accuracy: 33 correct gives 0.235714285714
compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy',
'liveness', 'instrumentalness', 'acousticness']

Identified 6 out of 7 labels

hunter5474	4	of 20	(0.200000)
jungleprince	14	of 27	(0.518519)
sana	6	of 13	(0.461538)
mlhopp	1	of 20	(0.050000)
1246241522	7	of 17	(0.411765)
connor	1	of 18	(0.055556)

Trail 2 Part 4b:

```
python run.py ann 1 false data/data_2_16_16.txt danceability  
energy liveness instrumentalness acousticness > ann2_4b.txt
```

Results:

Accuracy: 42 correct gives 0.28 compared
to guessing 0.142857142857

Best feature set: ['danceability', 'energy',
'liveness', 'instrumentalness', 'acousticness']

Identified 3 out of 7 labels

jungleprince	7	of 22	(0.318182)
1246241522	34	of 36	(0.944444)
naomimusgrave	1	of 28	(0.035714)

Trail 2 Part 4c:

```
python run.py ann 1 false data/data_2_19_16.txt danceability  
energy liveness instrumentalness acousticness > ann2_4c.txt
```

Results:

Accuracy: 39 correct gives 0.26 compared
to guessing 0.142857142857

Best feature set: ['danceability', 'energy',
'liveness', 'instrumentalness', 'acousticness']

Identified 3 out of 7 labels

jungleprince	4	of 23	(0.173913)
1246241522	34	of 34	(1.000000)
naomimusgrave	1	of 27	(0.037037)

Trail 2 Part 4d:

```
python run.py ann 1 false data/data_2_8_16.txt danceability  
energy liveness instrumentalness acousticness > ann2_4d.txt
```

Results:

Accuracy: 19 correct gives 0.316666666667
compared to guessing 0.333333333333

Best feature set: ['danceability', 'energy',
'liveness', 'instrumentalness', 'acousticness']

Identified 3 out of 3 labels

mlhopp	1	of 26	(0.038462)
naomimusgrave	14	of 15	(0.933333)
connor	4	of 19	(0.210526)

Trail 3 Part 1:

python run.py ann 1 true data/data_2_16_16.txt > ann3_1.txt

Results:

Accuracy: 47 correct gives 0.313333333333
compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy',
'instrumentalness', 'acousticness']

Identified 5 out of 7 labels

jungleprince	10	of 24	(0.416667)
mlhopp	1	of 20	(0.050000)
connor	1	of 19	(0.052632)
1246241522	34	of 34	(1.000000)
naomimusgrave	1	of 32	(0.031250)

Trail 3 Part 2:

python run.py ann 1 false data/data_2_16_16.txt > ann3_2.txt

Results:

Accuracy: 44 correct gives 0.293333333333 compared to
guessing 0.142857142857

Best feature set: ['danceability', 'energy', 'liveness',
'loudness', 'speechiness', 'tempo', 'valence',
'instrumentalness', 'acousticness']

Identified 2 out of 7 labels

jungleprince	8	of 20	(0.400000)
1246241522	36	of 36	(1.000000)

Trail 3 Part 3:

```
python run.py ann 1 false data/data_2_16_16.txt danceability
energy instrumentalness acousticness > ann3_3.txt
```

Results:

Accuracy: 50 correct gives 0.333333333333 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy', 'instrumentalness', 'acousticness']

Identified 3 out of 7 labels

jungleprince	7	of 14	(0.500000)
1246241522	42	of 42	(1.000000)
naomimusgrave	1	of 25	(0.040000)

Trail 3 Part 4a:

```
python run.py ann 1 false data/data_3_1_16.txt danceability
energy instrumentalness acousticness > ann3_4a.txt
```

Results:

Accuracy: 55 correct gives 0.243362831858 compared to guessing 0.0909090909091

Best feature set: ['danceability', 'energy', 'instrumentalness', 'acousticness']

Identified 2 out of 11 labels

jungleprince	17	of 30	(0.566667)
1246241522	38	of 40	(0.950000)

Trail 3 Part 4b:

```
python run.py ann 1 false data/data_3_8_16.txt danceability
energy instrumentalness acousticness > ann3_4b.txt
```

Results:

Accuracy: 25 correct gives 0.178571428571 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy', 'instrumentalness', 'acousticness']

Identified 5 out of 7 labels

jungleprince	13	of 20	(0.650000)
mlhopp	1	of 21	(0.047619)
connor	2	of 22	(0.090909)
naomimusgrave	3	of 23	(0.130435)
sana	6	of 8	(0.750000)

Trail 3 Part 4c:

```
python run.py ann 1 false data/data_2_19_16.txt danceability  
energy instrumentalness acousticness > ann3_4c.txt
```

Results:

Accuracy: 41 correct gives 0.273333333333
compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy',
'instrumentalness', 'acousticness']

Identified 2 out of 7 labels

jungleprince	7	of 21	(0.333333)
1246241522	34	of 34	(1.000000)

Trail 3 Part 4d:

```
python run.py ann 1 false data/data_2_8_16.txt danceability  
energy instrumentalness acousticness > ann3_4d.txt
```

Results:

Accuracy: 28 correct gives 0.466666666667
compared to guessing 0.333333333333

Best feature set: ['danceability', 'energy',
'instrumentalness', 'acousticness']

Identified 3 out of 3 labels

mlhopp	4	of 20	(0.200000)
naomimusgrave	19	of 21	(0.904762)
connor	5	of 19	(0.263158)

Trail 4 Part 1:

```
python run.py ann 1 true data/data_2_19_16.txt > ann4_1.txt
```

Results:

Accuracy: 45 correct gives 0.3 compared to guessing
0.142857142857

Best feature set: ['danceability', 'energy', 'liveness',
'instrumentalness', 'acousticness']

Identified 4 out of 7 labels

jungleprince	7	of 21	(0.333333)
1246241522	35	of 35	(1.000000)
naomimusgrave	2	of 36	(0.055556)
connor	1	of 17	(0.058824)

Trail 4 Part 2:

```
python run.py ann 1 false data/data_2_19_16.txt > ann4_2.txt
```

Results:

Accuracy: 46 correct gives 0.3066666666667 compared to guessing 0.1666666666667

Best feature set: ['danceability', 'energy', 'liveness', 'loudness', 'speechiness', 'tempo', 'valence', 'instrumentalness', 'acousticness']

Identified 3 out of 6 labels

jungleprince	4	of 17	(0.235294)
1246241522	40	of 41	(0.975610)
naomimusgrave	2	of 34	(0.058824)

Trail 4 Part 3:

```
python run.py ann 1 false data/data_2_19_16.txt > ann4_3.txt
```

Results:

Accuracy: 38 correct gives 0.2533333333333 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy', 'liveness', 'instrumentalness', 'acousticness']

Identified 2 out of 7 labels

jungleprince	8	of 21	(0.380952)
1246241522	30	of 31	(0.967742)

Trail 4 Part 4a:

```
python run.py ann 1 false data/data_2_8_16.txt danceability  
energy liveness instrumentalness acousticness > ann4_4a.txt
```

Results:

Accuracy: 24 correct gives 0.4 compared to guessing 0.3333333333333

Best feature set: ['danceability', 'energy', 'liveness', 'instrumentalness', 'acousticness']

Identified 3 out of 3 labels

mlhopp	6	of 16	(0.375000)
naomimusgrave	14	of 24	(0.583333)
connor	4	of 20	(0.200000)

Trail 4 Part 4b:

```
python run.py ann 1 false data/data_2_16_16.txt danceability  
energy liveness instrumentalness acousticness > ann4_4b.txt
```

Results:

Accuracy: 38 correct gives 0.253333333333 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy', 'liveness', 'instrumentalness', 'acousticness']

Identified 2 out of 7 labels

jungleprince	6	of 17	(0.352941)
1246241522	32	of 34	(0.941176)

Trail 4 Part 4c:

```
python run.py ann 1 false data/data_3_1_16.txt danceability  
energy liveness instrumentalness acousticness > ann4_4c.txt
```

Results:

Accuracy: 49 correct gives 0.216814159292 compared to guessing 0.0909090909091

Best feature set: ['danceability', 'energy', 'liveness', 'instrumentalness', 'acousticness']

Identified 3 out of 11 labels

jungleprince	12	of 25	(0.480000)
1246241522	36	of 37	(0.972973)
naomimusgrave	1	of 22	(0.045455)

Trail 4 Part 4d:

```
python run.py ann 1 false data/data_3_8_16.txt danceability  
energy liveness instrumentalness acousticness > ann4_4d.txt
```

Results:

Accuracy: 32 correct gives 0.228571428571 compared to guessing 0.142857142857

Best feature set: ['danceability', 'energy', 'liveness', 'instrumentalness', 'acousticness']

Identified 6 out of 7 labels

naomimusgrave	6	of 18	(0.333333)
hunter5474	3	of 22	(0.136364)
jungleprince	9	of 18	(0.500000)
sana	7	of 21	(0.333333)
1246241522	4	of 18	(0.222222)
connor	3	of 18	(0.166667)

Trail 5 Part 1:

```
python run.py ann 1 true data/data_2_8_16.txt > ann5_1.txt
```

Results:

Accuracy: 32 correct gives 0.533333333333 compared to guessing 0.333333333333

Best feature set: ['energy', 'liveness', 'speechiness', 'tempo', 'acousticness']

Identified 3 out of 3 labels

mlhopp	12	of 17	(0.705882)
naomimusgrave	18	of 19	(0.947368)
connor	2	of 24	(0.083333)

Trail 5 Part 2:

```
python run.py ann 1 false data/data_2_8_16.txt > ann5_2.txt
```

Results:

Accuracy: 20 correct gives 0.333333333333 compared to guessing 0.333333333333

Best feature set: ['danceability', 'energy', 'liveness', 'loudness', 'speechiness', 'tempo', 'valence', 'instrumentalness', 'acousticness']

Identified 2 out of 3 labels

mlhopp	11	of 14	(0.785714)
naomimusgrave	9	of 22	(0.409091)

Trail 5 Part 3:

```
python run.py ann 1 false data/data_2_8_16.txt > ann5_3.txt
```

Results:

Accuracy: 18 correct gives 0.3 compared to guessing 0.333333333333

Best feature set: ['energy', 'liveness', 'speechiness', 'tempo', 'acousticness']

Identified 3 out of 3 labels

mlhopp	2	of 19	(0.105263)
naomimusgrave	12	of 19	(0.631579)
connor	4	of 22	(0.181818)

Trail 5 Part 4a:

```
python run.py ann 1 false data/data_2_19_16.txt energy liveness  
speechiness tempo acousticness > ann5_4a.txt
```

Results:

Accuracy: 30 correct gives 0.2 compared to guessing
0.142857142857
Best feature set: ['energy', 'liveness', 'speechiness',
'tempo', 'acousticness']
Identified 1 out of 7 labels
1246241522 30 of 30 (1.000000)

Trail 5 Part 4b:

```
python run.py ann 1 false data/data_2_16_16.txt energy liveness  
speechiness tempo acousticness > ann5_4b.txt
```

Results:

Accuracy: 29 correct gives 0.193333333333 compared to
guessing 0.142857142857
Best feature set: ['energy', 'liveness', 'speechiness',
'tempo', 'acousticness']
Identified 2 out of 7 labels
1246241522 27 of 30 (0.900000)
naomimusgrave 2 of 24 (0.083333)

Trail 5 Part 4c:

```
python run.py ann 1 false data/data_3_1_16.txt energy liveness  
speechiness tempo acousticness > ann5_4c.txt
```

Results:

Accuracy: 29 correct gives 0.128318584071 compared to
guessing 0.0909090909091
Best feature set: ['energy', 'liveness', 'speechiness',
'tempo', 'acousticness']
Identified 2 out of 11 labels
jungleprince 2 of 36 (0.055556)
1246241522 27 of 28 (0.964286)

Trail 5 Part 4d:

```
python run.py ann 1 false data/data_3_8_16.txt energy liveness  
speechiness tempo acousticness > ann5_4d.txt
```

Results:

Accuracy: 25 correct gives 0.178571428571 compared to
guessing 0.142857142857

Best feature set: ['energy', 'liveness', 'speechiness',
'tempo', 'acousticness']

Identified 4 out of 7 labels

jungleprince	14	of 16	(0.875000)
mlhopp	3	of 17	(0.176471)
hunter5474	2	of 20	(0.100000)
connor	6	of 16	(0.375000)

—