

Predicting a song's genre and popularity based on it's audio features

CMSE 202 - Section 002

by Amelia, Christian, Ricky, Shria, Julia, Laura

Goal of the Project

Create a model that predicts a song's:

Genre and Popularity

Using various audio features from the Spotify API:

Beats per Minute, Energy, Danceability, Loudness, Liveness, Valence, Length, Acousticness, Speechiness

Index	Title	Artist	Top Genre	Year	Beats Per Minute (BPM)	Energy	Danceability	Loudness (dB)	Liveness	Valence	Length (Duration)	Acousticness	Speechiness	Popularity
1	Sunrise	Norah Jones	adult standards	2004	157	30	53	-14	11	68	201	94	3	71
2	Black Night	Deep Purple	album rock	2000	135	79	50	-11	17	81	207	17	7	39
3	Clint Eastwood	Gorillaz	alternative hip hop	2001	168	69	66	-9	7	52	341	2	17	69
4	The Pretender	Foo Fighters	alternative metal	2007	173	96	43	-4	3	37	269	0	4	76
5	Waitin' On A Sunny Day	Bruce Springsteen	classic rock	2002	106	82	58	-5	10	87	256	1	3	59
...
1990	Heartbreak Hotel	Elvis Presley	adult standards	1958	94	21	70	-12	11	72	128	84	7	63
1991	Hound Dog	Elvis Presley	adult standards	1958	175	76	36	-8	76	95	136	73	6	69
1992	Johnny B. Goode	Chuck Berry	blues rock	1959	168	80	53	-9	31	97	162	74	7	74
1993	Take Five	The Dave Brubeck Quartet	bebop	1959	174	26	45	-13	7	60	324	54	4	65

Raw Data

1994 rows x 15 columns

Fixing Genres

```
genres = ['country', 'dance', 'electro', 'folk']

for i in range(len(df['Top Genre'])):
    genre = df['Top Genre'][i].lower()
    for g in genres:
        if g in genre:
            df['Top Genre'][i] = g
            break
    else:
        df['Top Genre'][i] = 'unknown'
```

Fixing Popularity

```
df['Popularity'] = np.where(df['Popularity'] < 70, 'Average', 'High')
```

Index	Title	Artist	Top Genre	Year	Beats Per Minute (BPM)	Energy	Danceability	Loudness (dB)	Liveness	Valence	Length (Duration)	Acousticness	Speechiness	Popularity
1	Sunrise	Norah Jones	unknown	2004	157	30	53	-14	11	68	201	94	3	High
2	Black Night	Deep Purple	rock	2000	135	79	50	-11	17	81	207	17	7	Average
3	Clint Eastwood	Gorillaz	hip hop	2001	168	69	66	-9	7	52	341	2	17	Average
4	The Pretender	Foo Fighters	metal	2007	173	96	43	-4	3	37	269	0	4	High
5	Waitin' On A Sunny Day	Bruce Springsteen	rock	2002	106	82	58	-5	10	87	256	1	3	Average
...
1990	Heartbreak Hotel	Elvis Presley	unknown	1958	94	21	70	-12	11	72	128	84	7	Average
1991	Hound Dog	Elvis Presley	unknown	1958	175	76	36	-8	76	95	136	73	6	Average
1992	Johnny B. Goode	Chuck Berry	rock	1959	168	80	53	-9	31	97	162	74	7	High
1993	Take Five	The Dave Brubeck Quartet	unknown	1959	174	26	45	-13	7	60	324	54	4	Average

Cleaned Data

Notice Top Genre and Popularity

X_data

	Beats Per Minute (BPM)	Energy	Danceability	Loudness (dB)	Liveness	Valence	Length (Duration)	Acousticness	Speechiness
0	157	30	53	-14	11	68	201.0	94	3
1	135	79	50	-11	17	81	207.0	17	7
2	168	69	66	-9	7	52	341.0	2	17
3	173	96	43	-4	3	37	269.0	0	4
4	106	82	58	-5	10	87	256.0	1	3
...
1989	94	21	70	-12	11	72	128.0	84	7
1990	175	76	36	-8	76	95	136.0	73	6
1991	168	80	53	-9	31	97	162.0	74	7
1992	174	26	45	-13	7	60	324.0	54	4
1993	133	50	49	-10	16	83	148.0	74	3

Popularity Y_data

0	High
1	Average
2	Average
3	High
4	Average
	...
1989	Average
1990	Average
1991	High
1992	Average
1993	Average

Name: Popularity,

Genre Y_data

0	unknown
1	rock
2	hip hop
3	metal
4	rock
	...
1989	unknown
1990	unknown
1991	rock
1992	unknown
1993	unknown

Name: Top Genre,

RandomForestClassifier



Popularity Classification

Popularity Classification:

Accuracy=0.7293233082706767, Precision=0.673234891113074, Recall=0.7293233082706767, F1 Score=0.6588925099482931

Cross-validation scores for Popularity Classification: [0.72681704 0.69172932 0.72932331 0.72431078 0.70854271]

Average cross-validation score for Popularity Classification: 0.72

Pop Confusion Matrix:

	Average	High
Average	280	12
High	96	11

Genre Classification

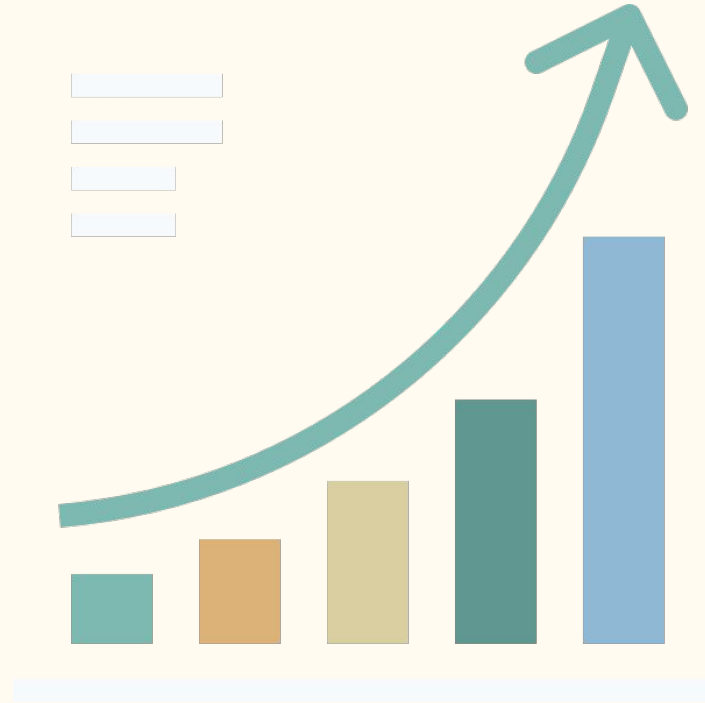
Genre Classification:
Accuracy=0.44611528822055135, Precision=0.3718125651820182, Recall=0.44611528822055135, F1 Score=0.3813399171135454

Cross-validation scores for Genre Classification: [0.39598997 0.38345865 0.43609023 0.41102757 0.38693467]
Average cross-validation score for Genre Classification: 0.40

Genre Confusion Matrix:

	country	dance	electro	folk	hip hop	metal	pop	rock	soul	unknown
country	0	0	0	0	0	0	0	1	0	1
dance	0	2	0	0	0	1	5	19	0	4
electro	0	1	0	0	0	0	0	1	0	0
folk	0	0	0	0	0	0	0	1	0	0
hip hop	0	0	0	0	2	0	0	0	0	2
metal	0	1	0	0	0	1	0	14	0	2
pop	0	3	0	0	0	0	5	32	0	21
rock	0	1	0	0	0	2	10	134	0	14
soul	0	2	0	0	0	0	2	7	0	3
unknown	0	2	0	0	0	0	5	63	0	34

How to improve the results?



Tying it all together

```
Enter a song name: Rocket Man  
Enter the artist of the song: Elton John  
Audio features for Rocket Man by Elton John:
```

```
Beats Per Minute (BPM): 136.571  
Energy: 0.532  
Danceability: 0.601  
Loudness (dB): -9.119  
Liveness: 0.0925  
Valence: 0.341  
Length (Duration): 281613  
Acousticness: 0.432  
Speechiness: 0.0286
```

```
Popularity Prediction: Average  
Spotify's Actual Popularity Rating (<70 we consider average popularity): 83  
Genre Prediction: Pop
```

Thank You

:)